Structural Analysis of Nucleation Mechanism in Curli Biogenesis using Nuclear Magnetic Resonance Spectroscopy

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Madhu Nagaraj

aus Bhadravathi, India

1. Referentin:	Prof. Dr. Christiane Ritter
2. Referent:	Prof. Dr. Michael Steinert
eingereicht am:	08.08.2012
mündliche Prüfung (Disputation) am:	05.11.2012

Drukjahr: 2013

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch die Mentorin der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge:

Ritter, C., Nagaraj, M., Zimmer, A.,: NMR Structural investigations of the curli subunits CsgA and CsgB (Poster) FASEB Summer Research Conference - The Basic Origins and Medical Consequences of Protein Aggregation, Snowmass Village, USA (2011)

Ritter, C., Nagaraj, M., Zimmer, A.,: NMR structural analysis of the nucleator-dependent biogenesis of the functional amyloid curli (Poster) International Bunsen Discussion Meeting on Structure of Amyloid Fibrils and Mechanism of Amyloid Formation, Halle/Saale (2011)

Eberth A., Nagaraj, M., Ritter C.: Structural investigations of the major curlin subunit CsgA – a functional amyloid (Poster) Jacques Monod Conferences Protein folds in Infectious and neurodegenerative diseases, Aussois, France (2009)

In memory of my parents

Acknowledgements

I would like to thank my mentor Prof.Dr. Christiane Ritter for her support, guidance and giving me the opportunity to work independently in her group. She has always been supportive and has helped me adapt to the German ways. Christiane's vast knowledge in NMR has helped to steer through the challenging parts of the project on curli amyloids. Her accommodative nature and encouragement have made my time here a truly comfortable yet enriching learning experience. I will always cherish the parties she threw. I am grateful to Thorsten Luehrs whose expertise with NMR has helped me achieve superior results. His willingness to troubleshoot the NMR instrument even at late nights has helped me a great deal.

I am ever so grateful to Dr.Alexander Eberth, a truly exceptional scientist and a friend for not only providing a scientifically stimulating atmosphere in the group but also for the fun times. Skiing on the Harz, and weekly football matches with Alex and Agnes will be a memory I will always cherish. Agnes and her capability to work hard regardless of the circumstances have and will always inspire me to do the same. I thank her for some brilliant discussions, translation of German documents, and of course all the snacks.

I thank all my thesis committee members Prof.Dr.Victor Wray, Prof.Dr.Dirk Heinz and Prof.Dr.Petra Dersch for providing great scientific discussions. I would like to thank Prof. Dr. Manfred Rohde for the excellent scanning electron microscopy picture which is presented in this work. I also thank him for providing me an opportunity to learn negative staining EM under his supervision, which helped me in my work presented here.

I would like to extend my gratitude to Jenny without whom my PhD would not have been as fun with all the super football matches and summer barbeques. On the same note, I thank Caro for helping with the protein purifications and for the friendly chats. Although I have known Dr. Mumdooh Ahmed only for a short duration he has been very friendly and a fun person I thank him sincerely for the help with NMR and also for the help with some pictures in my thesis. I thank all the other lab members including Puwei Yuan, Lichun He, Johannes Spehr and Tobias Schubeis for a friendly environment in the lab. I would also like to thank all the SB members and my Indian friends for providing a friendly atmosphere at HZI.

Last but not the least; I would like to thank Ramya for helping me a lot with my thesis writing.

I would like to thank my family for their love, encouragement and constant support.

Contents

Abbreviations

aa	Amino acid
Å	Ångström
AFM	Atomic force microscopy
AP	Alkaline phosphatise
App	Appendix
APS	Ammonium persulphate
BMRB	Biological magnetic resonance data bank
Bis-Tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methan
CD	Circular dichroism
CR	Congo red
CsgX	Genproduct of curli specific gene X
CV	Column volume
kDa	Kilo Dalton
E. coli	*Escherichia coli*
Ec	*Escherichia coli*
ECL	Enhanced Chemiluminescence
EM	Electron microscopy
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
FA	Formic acid
H/D	Hydrogen/Deuterium
HDX	Hydrogen/Deuterium Exchange
hNOE	Heteronuclear Overhauser effect
HSQC	Heteronuclear single quantum coherence
FT-IR	Fourier transform infrared
HRP	Horse radish peroxidase
IB	Inclusion body
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-β-D-thiolgalactoside
KPi	KH_2PO_4/K_2HPO_4

MHC	Major Histocompatibility complex
MWCO	Molecular weight cut off
Ni-NTA	Nickel-nitrilotriacetic acid-agarose
NOE	Nuclear Overhauser Effect
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser enhancement spectroscopy
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamid gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
R	Repeat
SDS	Sodium dodecyl sulphate
Sec	Secretory
TAE	Tris-acetate-EDTA
TBS	Tris buffered saline
TCA	Trichloroacetic acid
TEM	Transmission electron microscopy
TEMED	N, N, N', N' – Tetramethylendiamin
TFA	Trifluoroacetic acid
ThT	Thioflavin T
Tris	Tris (hydroxymethyl) aminomethane
UTI	Urinary tract infection
wt	Wild type

Amino acids are abbreviated using 1 or 3 letter code.

Latinisms are written in italic.

Summary

Amyloid fibrils are a class of protein structures that exhibit a common behavior and assemble into highly ordered aggregates. Amyloid fibrils, are much-studied phenomenon because of its implications for human health, since it is associated with many neurodegenerative diseases including Alzheimer's disease, Parkinson's disease and Huntington's disease (2006; Abrahams and Bax, 1991; Chiti and Dobson, 2006b). Recently an increasing number of proteins with no link to protein deposition diseases have been found in several organisms ranging from bacteria, fungi and insects to humans where they perform beneficial functions *in vivo* as native proteins known as functional amyloids. These fibrillar aggregates have the morphological, structural, and tinctorial properties that allow them to be classified as amyloid fibrils. One particularly well-studied form that expresses extracellular proteinaceous fibrils called curli, are used by *Escherichia coli* to colonize inert surfaces and mediate binding to host proteins.

The abbrerant protein folding giving rise to uncontrolled formation of amyloid fibers is the hallmark of diseased amyloids. In contrast to disease-associated amyloids, E. *coli* exhibits highly regulated amyloid formation which assembles curli fibrils on the cell surface using an elegant biogenesis machine. The study of curli assembly machinery provides a unique template for understanding controlled amyloid propagation. *In vivo*, amyloidogenesis of the major curli subunit protein CsgA is dependent on the minor curli subunit protein CsgB, which forms a dedicated nucleator for amyloid growth (Rochet and Lansbury, 2000a). Nucleation forms the rate-limiting step of amyloid propagation and its nature remains poorly understood.

In the present study, CsgB-wt was recombinantly produced, purified and fibrillised for hydrogen/deuterium exchange NMR experiments to determine sequence specific secondary structure. In conjunction with the studies on nucleation properties of the CsgB truncation mutants, we aimed to determine the structure function relationship of CsgB.

Mainly, the study comprised of detailed structural analysis of nucleator-dependent curli biogenesis using NMR. The secondary structure of CsgB revealed a repetitive structural pattern with four repeat sequences each forming a strand-turn-strand motif in the amyloid core. The N-terminal region showed an additional single beta strand and the C-terminal region displayed highly heterogeneous behaviour indicating its partial incorporation in the amyloid core. Mutational studies by truncation of both the N and C terminal region of

CsgB followed by analysis on its nucleation property with CsgA revealed that the N-terminal of CsgB was involved in sequence selective interface for seeding CsgA. The aforementioned study also indicated that the N- terminus of CsgB was involved in providing directionality of fiber growth during curli biogenesis. The C-terminus was involved in the self-folding propensity of CsgB which we could show to be the major rate limiting step during the nucleation of CsgA.

These results present the first sequence-specific data of structural elements of CsgB fibrils and provide a closer picture of the nucleation process at a structural level in curli biogenesis.

1. Introduction

1.1 Protein folding

Proteins are considered the building blocks of cellular systems. They are crucial to all living systems due to their involvement in all aspects of cell life including immune protection, cell-cell communication, physical support (hair and skin), and muscle movement (Stryer, 1966). Proteins and peptides are built from smaller individual amino acid units with each amino acid containing an α-carbon bonded to an α-amino group, an α-carboxyl group, one hydrogen, and a side chain, R, of varying lengths and functionalities (Stryer, 1966). The growth of a protein occurs by a condensation reaction between the free amine and carboxylic group of the different amino acids to form amide bonds (Stryer, 1966).

Proteins are usually composed of 20 different kinds of amino acids which spontaneously fold into a stable structure to be functional. Protein folding is a process of acquiring a unique, stable and kinetically accessible native structure with minimum free energy. Extensive work to define protein folding mechanisms has been carried out but we are still far from fully understanding the mechanism of protein folding. Structural prediction of a protein from its sequence is still challenging, hence understanding protein-folding mechanisms are much in demand for elucidating sequence-structure-function relationships (Ivankov and Finkelstein, 2004).

One of the most widely accepted concepts describing protein folding is the concept of the energy landscape (Chan and Dill, 1998; Dill and Chan, 1997), where folding properties of proteins are depicted by a funnel-like energy profile with a rough surface defined by thermodynamic and kinetic properties (**Figure 1**). The Funnel-like energy profile describes protein folding with the native state at the lowest minimum of the funnel, as the most stable configuration of the polypeptide chain. However, there are some known cases of inconsistencies where the native states of the proteins do not have the lowest minimum. One such case is that of amyloid fibrils, the formation which is of high significance.

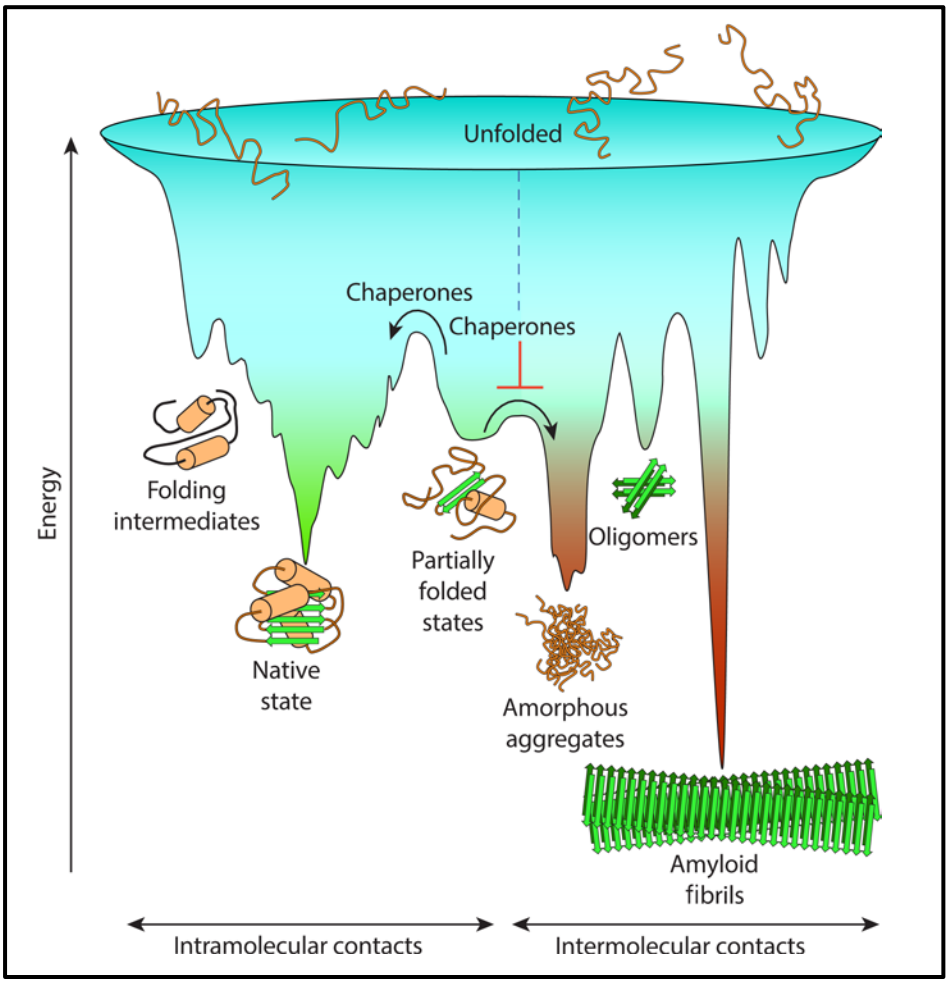


Figure 1: Funnel-shaped free-energy landscape of protein folding and aggregation

(Hartl et al., 2011).The green surface shows the multitude of conformations 'funneling' to the native state via intramolecular contacts *in vivo*, these steps may be accelerated by chaperones. Brown area shows the conformations moving toward amorphous aggregates or amyloid fibrils via intermolecular contacts. Destabilization of the native protein into partially folded form or intermediate protein folds formed during the folding process may result in aggregate formation which is normally prevented by molecular chaperones.

Protein folding reactions are highly complex and depend on several weak, non-covalent interactions allowing for different possible conformations that a protein can adopt. For example, in case of soluble proteins, hydrophobic forces play an important role in driving internalization of non-polar amino-acid residues (Hartl and Hayer-Hartl, 2009).

The sequestering of the hydrophobic side chains of the polar groups in combination with formation of the hydrogen bonds allows the protein to achieve the lowest free energy state (Stryer, 1966). The secondary structure of a protein depends on the types of amino acids present within its protein sequence (Stryer, 1966). The most common secondary structures are alpha helices and beta sheets. The secondary structure of the α-helix is formed by the

hydrogen bonding between the backbone NH and CO groups of the same strand whereas the β-strand forms hydrogen bonding with other β-strands (**Figure 2**) (Stryer, 1966).These β-strands thus stabilize each other through intermolecular hydrogen bonding forming β-sheets (West et al., 1999). Tertiary structures build on these secondary structures to form other higher-order structures such as β-sheets, βαβ units, β-hairpins, and α-helix bundles (Stryer, 1966). Many weak bonds, including hydrogen bonds and Van Der Waals interactions, are formed in the protein-folding process. Hence hydrophobic interactions in combination with the above mentioned interactions make the overall free energy negative, stabilizing the protein structure (Berg JM, 2002).

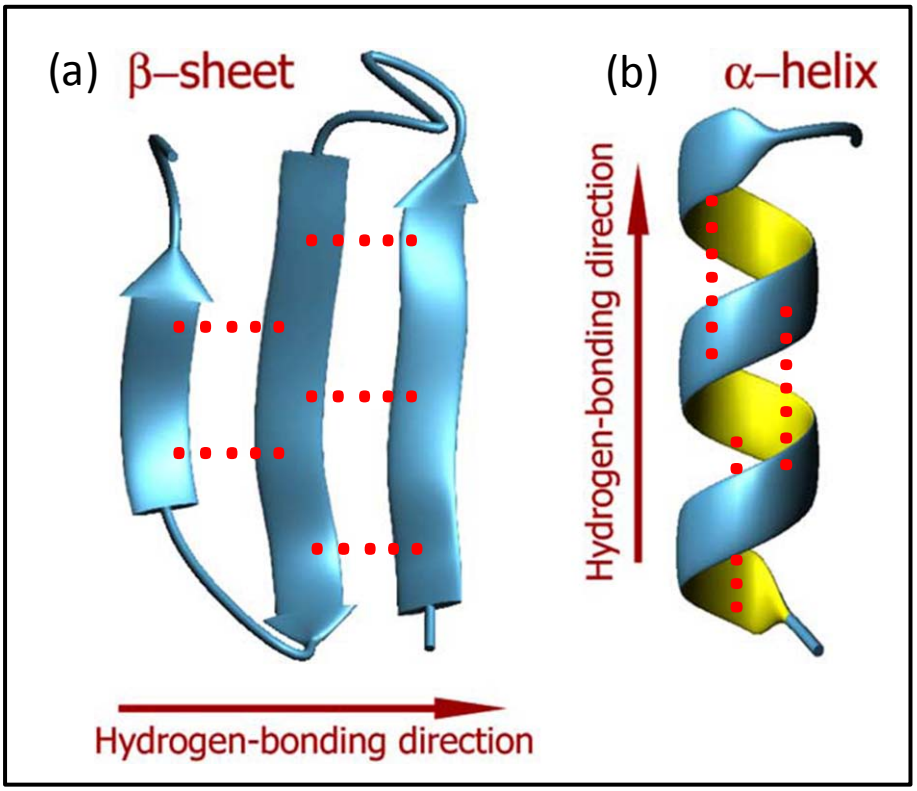


Figure 2: Secondary Structures of Proteins.

(a) β-sheet. **(b)** α-helix. The peptide backbones (blue line) can intramolecular hydrogen bond (red dotted line) as seen in the α-helix or intermolecular hydrogen bond (red dotted line) as seen in the β-sheet.

1.2 The Driving Force behind Protein Aggregation: Hydrophobic Clustering

Hydrophobic clustering or collapse is not caused by mere chance encounters of hydrophobic domains, but is driven by the ordering of solvent (water) upon these domains (Berg JM, 2002; Stryer, 1966). Water as a solid (ice) is more ordered than liquid water and thus has lower entropy. When a hydrophobic domain on a protein or peptide is placed in water, the water molecules are disturbed in its immediate vicinity to form a solvation shell. However, the presence of other protein molecules results in hydrophobic clustering of the protein molecules disfavoring the higher entropy solvation shell between protein and water molecules. Hence higher entropy provides a thermodynamically favorable event for protein aggregation (Berg JM, 2002; Stryer, 1966).

2. Amyloid

2.1 Amyloid History

The term ‘Amyloid’ was first coined by Rudolph Vichow in 1854 to describe the abnormal substance in brain tissue stained positively by iodine (Sipe and Cohen, 2000).

Amyloid was derived from the Latin amylum and the Greek amylon since the chemical nature of amyloid was thought to be starch-like because it stained with iodine (Sipe and Cohen, 2000). Later, amyloid was demonstrated to be protein aggregates with ordered structures (Bonar et al., 1969; Glenner et al., 1969; Glenner et al., 1971a; Glenner et al., 1971b; Kim et al., 1969; Shirahama and Cohen, 1967). Amyloids are distinct proteinaceous fibers traditionally associated with many human ailments such as neurodegenerative diseases.

Amyloids are insoluble fibrous protein aggregates possessing specific structural traits. They display a repetitive intermolecular cross-beta sheet motif unique among protein folds. This motif is a pattern characterized by repetitive arrays of beta-sheets that are parallel to the fibril axis, with their strands perpendicular to the axis. Once the amyloidogenic protein forms a potent nucleus having a propensity to sequester the other monomeric soluble proteins, this translates into a repetitive oligomer cooperatively.

2.2 **Amyloid in human disease**

Amyloids are generally viewed as an outcome of protein misfolding and aggregation. Amyloids are implicated in many chronic human diseases. Protein can fail to fold and deviate from the normal behavior due to many reasons. One of them are mutations which change the amino acid in the polypeptide chain and can lead the protein either to misfold or lose the propensity to fold causing loss of function. Several cases of inherited mutation cause disease in humans, for example cystic fibrosis or amyloid diseases such as early onset Alzheimer's, Huntington and sickle cell anemia. Another reason for the protein misfolding can occur due to mistakes in the protein translation machinery, where ribosomes can make mistakes in as many as 1 in every 20,000 amino acids, although levels can be higher or lower depending on the conditions (Kramer and Farabaugh, 2007; Kurland, 1992). These mistakes once again results in the protein misfolding. This occurs constantly in the cell and is typically taken care of by chaperones or quality control mechanisms resulting in protein degradation.

Apart from mutations, even changes in environmental conditions in and out of the cell, for example acidity and temperature or changes different from the normal physiological conditions to which the organism is accustomed to can also lead to errors in protein folding. (**Figure 3**)

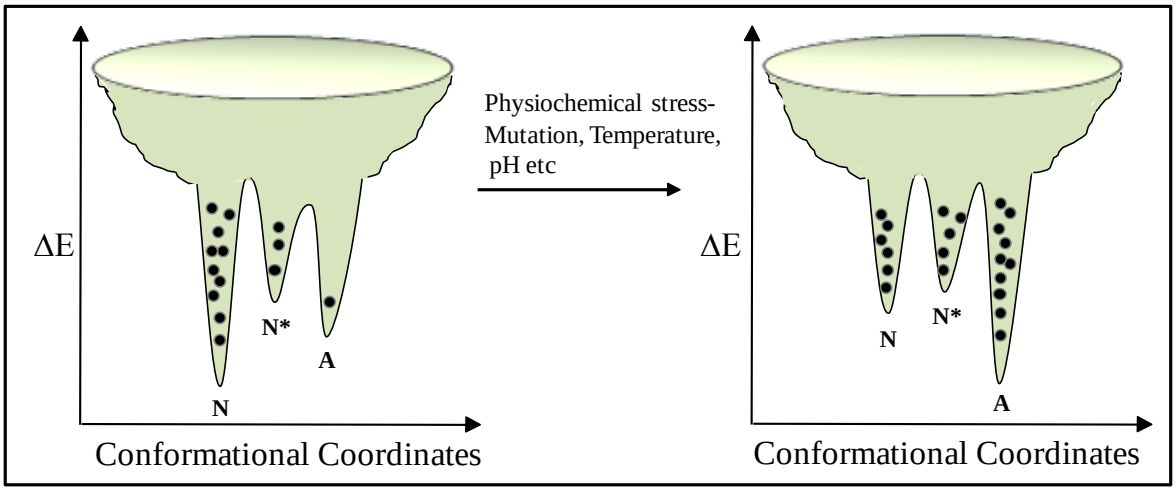


Figure 3: **Physiological stress in Amyloid formation**

N denotes native state; **N*** denotes native -like aggregation - competent state, and **A** denotes the aggregated state of the molecule. Different stresses may change the barriers between these states differently and may lead to conformer population shifts differently among the three states. (Agrawal et al., 2011)

This can result in proteins that are misfolded and therefore resistant to chaperoning and disposal mechanisms. Accumulations of this incorrectly folded species above critical concentration can lead to pathogenic fibrillization and thus make the protein more liable to become amyloid. Anomalous buildup and deposition of amyloid fibrils in organs may lead to amyloidosis, which is also a hallmark of various neurodegenerative disorders (**Table 1**). At least 18 proteins and polypeptides have been identified (Ramirez-Alvarado et al., 2000) which erroneously interact with one another or other cell components forming insoluble fibrils. Amyloids have been implicated in the pathology of more than 20 serious human diseases (Kelly, 1998). The general morphology and biophysical properties of all amyloid fibrils are extremely similar but show no correlation to the sequence homology or related native structure of the precursor proteins involved in various diseases.

Table 1: Amyloidoses - human disease associated with amyloid fibril formation
(Reviewed by (Chiti and Dobson, 2006b; Herczenik and Gebbink, 2008).

Disease	Associated proteins	Native structure	Affected tissue
Neurodegenerative diseases			
Spongiform encephalopathies	Prion protein	Unfolded N-term. α-helical C-term.	Brain, peripheral nervous system
Familiar amyloidic polyneuropathy	Transthyretin	All-β, prealbumin like	Peripheral nervous system
Alzheimer´s disease	Amyloid β, tau	Unfolded	Brain
Huntington´s disease	Huntingtin	Largely unfolded	Brain
Parkinson´s disease	α-Synuclein	Unfolded	Brain
Lewy-body dementia	α-Synuclein	Unfolded	Brain
Amyotrophic lateral sclerosis	Superoxide dismutase I	All-β, Ig like	Brain
Nonneuropathic systemic amyloidoses			
Hemodialysis-related amyloidosis	β_2-Microglobulin	All-β, Ig like	Osteoarticular tissues
Senile systemic amyloidosis	Transthyretin	All-β, prealbumin like	Microvasculature
Primary systemic amyloidosis	Immunoglobulin light chain	All-β, Ig like	Most tissues
Secondary (reactive) systemic amyloidosis	Serum amyloid A	All-α, unknown fold	Most tissues
Nonneuropathic localised amyloidoses			
Familial subepithelial corneal amyloidosis	Lactoferrin	α+ β, periplasmic-binding protein like	Cornea
Cataract	Crystallin	All-β, crystalline-like	Eye

Disease	Associated proteins	Native structure	Affected tissue
Type II diabetes	Amylin (islet amyloid polypeptide IAPP)	Unfolded	Pancreas (islet)
Injection localised amyloidosis	Insulin	All-α, insulin-like	Skin, muscles

Until recently, amyloids were thus considered "biological blunders" associated with protein misfolding. Amyloids formed large aggregates sticking to each other to form protease resistant clumps in the cell which are very hard to clear causing cell death especially during the absence of specific protective mechanism.

The cell has several mechanisms to protect against protein aggregation. Proteins initially synthesized and folded may have there amyloidogenic patches exposed. So other molecules, namely chaperone proteins, assist the proteins fold correctly and avoid the amyloid state. Many more cellular mechanisms recognize and destroy amyloids that have been formed and proteins that are susceptible may be sequestered in their own compartments. Approximately there are hundreds of genes that help to protect against the amyloid threat. Other than these protective mechanisms, the proteins themselves tend to have their amyloidogenic patches safely hidden away in the protein interior (Schnabel, 2010).

Table 2: Proteins that form functional amyloid fibrils

(Reviewed by (Chiti and Dobson, 2006b; Fowler et al., 2007b; Greenwald and Riek, 2010; Otzen and Nielsen, 2008).

Protein	Species	Function
	Bacterial amyloids	
Curli, Tafi	*E. coli* and other *Enterobacteriaceae*	Adhesion, cell invasion, biofilm formation, binding to host proteins
FapC	*Pseudomonas*	Biofilm formation
MTP	*Mycobacterium tuberculosis*	Adherence factor
Microcin E492	*Klebselia pneumonia*	Bacteriocin, membrane pore-forming, amyloid form inactive
Chaplins	*Streptomycetes coelicolor* and other *Streptomyces*	Assisting aerial hyphae formation via modulation of water surface tension
Harpins	*Xanthomonas* and other plant pathogenic bacteria	Destabilisation of plant membranes, induction of cell death
TasA	*Bacillus.Subtilis*	Biofilm
	Eukaryotic amyloids	
Pmel17	Mammals	Scaffolding and Sequestration of toxic intermediates during melanin synthesis

Protein	Species	Function
Peptide hormones	Mammals	Sorting, storage and release of hormones
Hydrophobins	Most fungi	Modulation of surface attachment and aerial hyphae formation
Chorion proteins	Insects and fish	Structure-bearing and protection of oocytes and developing embryo
Spidroins	*Nephila clavipes* (spider)	Structure-bearing, formation of spider silk
	Prion proteins	
ECTO-NOX	Mammals	[NADP(H)] oxidation and protein disulfide-thiol interchange
Mot3	*Saccharomyces cerevisiae*	Regulation of transcription of cell wall remodelling genes, prion form inactive
Swi1p	*Saccharomyces cerevisiae*	Chromatin remodelling
Rnq1p	*Saccharomyces cerevisiae*	Enhancing the inducibility of other prions
Ure2p	*Saccharomyces cerevisiae*	Regulation of nitrogen catabolism, prion form inactive
Sup35	*Saccharomyces cerevisiae*	Regulation of translation termination, prion form inactive
CPEB	*Aplysia* (sea slug)	Memory storage, leading to long term synaptic changes
HET-s	*Podospora anserina*	Regulation of heterokaryon formation

2.3 **Functional Amyloids**

However, not all amyloids are pathogenic. There is a vast body of evidence confirming the beneficial functions of amyloid structures in nature as functional amyloids .In contrast to disease-associated amyloids, functional amyloids are the product of coordinated and regulated cellular processes that ensure that amyloidogenesis does not result in cell damage and death (Fowler et al., 2007a; Hammer et al., 2008). Although functional amyloids were first discovered in microbes, their presence and relevance in higher organisms including humans has been identified (Aguzzi, 2008; Fowler et al., 2006; True and Lindquist, 2000) (**Table 2**). Functional amyloids are indespensible for important physiological functions in the cell and provide insight into the mechanisms of protein homeostasis, folding, and misfolding.

2.4 Unique - Functional Amyloid

Amyloids are attractive building materials because its robust design and simple repetitive structure make for very durable and therefore potentially metabolically cheap material (Otzen, 2010). But this requires that the production of amyloid be put under firm control. The robustly stacked β-strands shows an estimate of one misalignment per 30,000 strands in a typical amyloid structure (Knowles et al., 2006). Many functional amyloid resists boiling SDS which has been used as a convenient selection criterion for their isolation (Dueholm et al., 2010; Jordal et al., 2009). Several cases are known where functional amyloid can be dissolved only in high concentrations of formic acid (Aguzzi, 2008; Dueholm et al., 2010; Jordal et al., 2009). Furthermore, functional amyloid has typically been hardwired to evolve into fibrils, implicating that this process is robust enough to occur over a broad range of conditions rather than a small "window of opportunity". For example Curli fibrillates to the same amyloid structure *in vitro* under a broad range of pH, temperature, concentration and ionic strength, according to fiber diffraction and Fourier Transform Infrared Spectroscopy (Dueholm et al., 2011). Despite that robustness even functional amyloids can misfold under unphysiological extremes of pH, for example fungal HET-s protein shows different structures at pH 3 versus pH 7, as seen by recent solid state NMR studies (Wasmer et al., 2008) and previously already by FTIR.

This evolutionarily optimized robustness makes functional amyloid as reliable building material such as naturally occurring barnacle adhesive plaques (Barlow et al., 2010) to biotechnological applications such as templates for gold nanowires (Scheibel et al., 2003) and biomimetic silks in materials and medical applications (Heim et al., 2009). The amyloid state also helps as storage state for peptide hormones normally secreted in secretory granules of the endocrine system, (Maji et al., 2009) though in this case extreme stability is not desired; rather, the amyloid here has to constitute a rapidly mobilizable source of material and therefore the amyloid can probably dissociate quite simply by dilution.

2.5 Possible mechanism of amyloid clearance

The robust durability of the amyloid raises the question of its recycling. The large quantities of functional amyloid produced would lead to accumulation of biomass if they could not be degraded properly in bacteria (**Table 2**). In case of functional amyloid, recent discovery shows evidence of recycling by secretion of high levels of proteases for example in *Candidatus Epiflobacter* species (*Saprospiraceae, Bacteroidetes*) (Xia et al., 2008) and in sulfur-oxidizing bacteria *Thiothrix* (Larsen et al., 2008; Otzen, 2010). Similarly, there has been evidence of recycling and rearranging in toxic amyloid for example in Alzheimer's peptide Abeta (Sanchez et al., 2011). However, the exact mechanisms of recycling are mostly not understood.

2.6 General Amyloid features

Amyloids have several finger print biochemical properties, mainly including binding and causing birefringence of the dye Congo red (CR) and a spectral shift of the dye Thioflavin T (ThT) (Ghiso and Frangione, 2002; Sipe and Cohen, 2000). They also exhibit characteristic cross–β sheet x-ray diffraction patterns (**Figure 4**).

Diseased associated amyloids are thought to assemble through a conserved pathway, with soluble monomers forming ordered oligomeric intermediate structures and finally, fibers. While in functional amyloid several pathways exits but formation of oligomeric intermediates are rare. In general the formation of a stable nucleus is believed to be the rate limiting step in the amyloid formation .This gives rise to a typical growth curve with lag, log and stationary phase that is well represented by a sigmoidal curve. The propensity of protein to aggregate under optimal conditions has been demonstrated, indicating the role of peptide backbone interactions in amyloid formation (Chiti et al., 1999) .

Amyloid Binding to Thioflavin T and Congo red

Most amyloids bind Congo red. The orientation of Congo red molecules on amyloid fibrils makes the dye dichroic and birefringent. Congo red gives a characteristic apple-green birefringence when examined a characteristic shift in absorbance maximum from about 490 nm to about 540 nm is seen (Howie and Brewer, 2009; Kim et al., 2003).

Another amyloid-specific dye is Thioflavin T, a benzothiazole salt. It was first discovered in 1959, the dye shows enhanced fluorescence upon binding to amyloid in tissue sections

(Vassar and Culling, 1959) and has since become a standard dye for amyloid detection. When it binds to beta sheet-rich structures present in amyloid aggregates, the dye displays enhanced fluorescence in combination with strong shift in its excitation wavelength (LeVine, 1999b). This change in fluorescent behavior of the dye can be due to the chemical interactions with the rigid, highly-ordered amyloid structure causing a change in the charge distribution of Thioflavin T (Wolfe et al., 2010). Detection is based on the appearance of a new excitation peak at 450 nm shift in the excitation maximum of roughly 100 nm, upon binding to amyloid fibrils that is responsible for enhanced fluorescence emission at 482 nm, a (LeVine, 1993; Naiki et al., 1989).

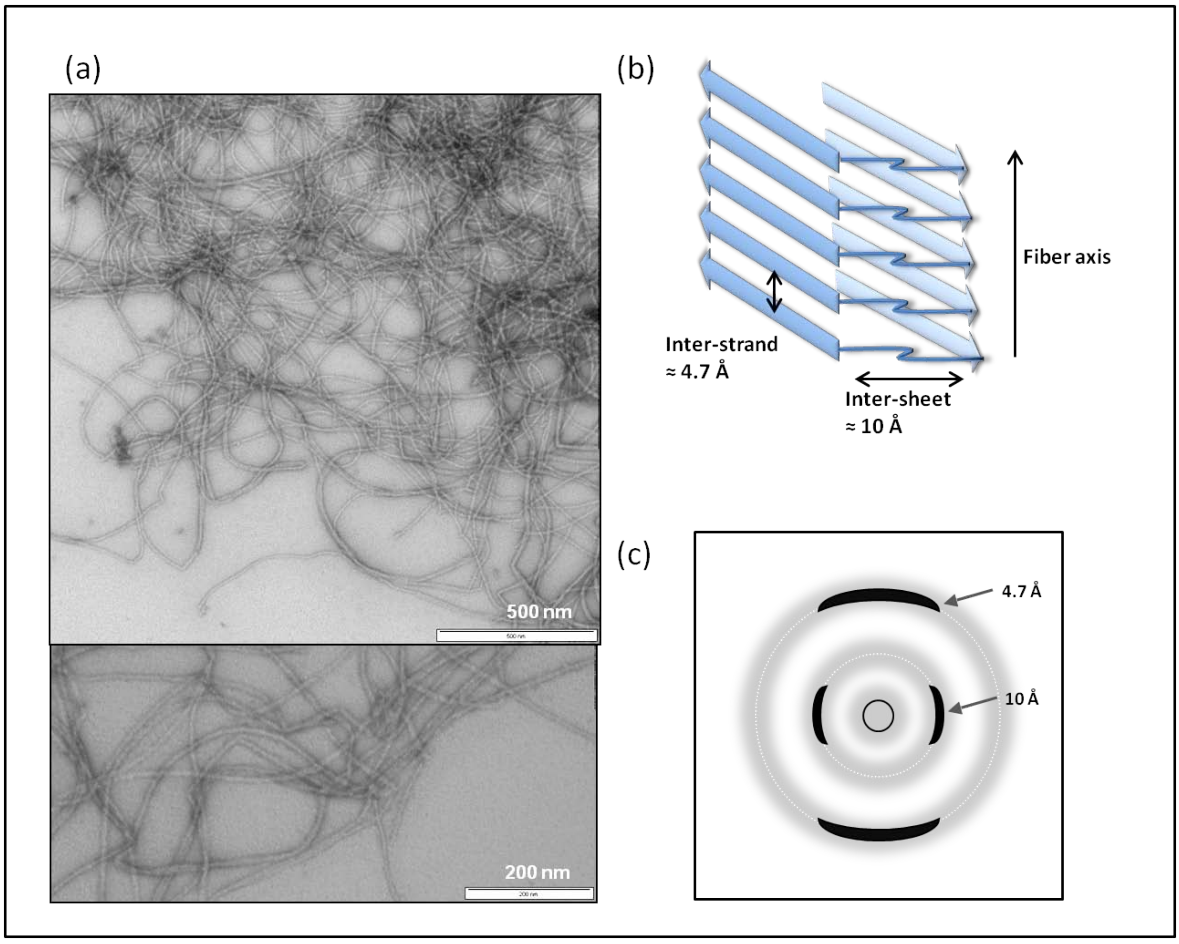


Figure 4: Fiber diffraction and TEM of amyloid fibers.

(a) A transmission electron micrograph of negatively stained amyloid fibers formed from E.coli Curli -CsgA. Magnification indicated on the right corner top picture (500nm) and bottom (200nm) **(b)** & **(c)** Depiction of X-ray fiber-diffraction pattern from amyloid fibrils exhibiting the characteristic reflections at 4.7 Å and≈10 Å. The meridional reflection at 4.7 Å results from the interstrand repeats, and the≈10 Å equatorial reflection arise from intersheet packing of cross β-sheet fold.

Morphology in Transmission Electron Microscopy

The amyloid fibrils are typically 5–15 nm in width and several micrometers in length. The fibrous components of amyloids as seen in electron microscopy by negative staining consist of multiple thin fibers, or protofilaments, that are entwined around each other. **(Figure 4a**)

Fiber diffraction pattern

X-ray fiber diffraction has been successfully used to examine the structure of the insoluble fibers. Fiber diffraction provides spatial information of frequently occurring distances in the structures. The fiber X-ray diffraction patterns of amyloid fibrils (Bonar et al., 1969) (Geddes et al., 1968) (Fig. 4c) display simple patterns with 4.7 to 4.8 Å meridional reflections and 10 Å equatorial reflections, arising from the molecular spacings between β-strands and between opposing β-sheets respectively. They are characteristic of a cross-β structure (Astbury et al., 1935; Pauling and Corey, 1951) in which the polypeptide chain is organized in β-sheets arranged parallel to the fibril axis with their constituent β-strands perpendicular to the fibril axis.

Amyloid formation

According to a popular model proposed by Lansbury, the transition from soluble monomeric proteins to polymeric and insoluble amyloid fibers is characterized by three distinct phases: a lag phase, a fiber elongation phase (growth) and a plateau or stationary phase, (**Figure 5**) that result in loss- or gain-of-function properties for the protein. In the lag phase, soluble protein assembles into a common intermediate or nucleus that is proposed to be toxic to membranes in pathogenic amyloids (Glabe and Kayed, 2006; Kayed et al., 2003). Monomers undergo conformational changes to form a potent nucleus which is followed by an elongation phase, in which the nuclei to sequester other monomeric soluble proteins translating into a repetitive oligomer and form larger polymers until saturation. The ‘nucleation phase', is thermodynamically unfavorable and requires more time, whereas ‘elongation phase', is much more favorable process and proceeds quickly. Thus, kinetics of amyloid formation is well represented by a sigmoidal curve. The formation of the intermediate is proposed to be the rate-limiting step of amyloidogenesis. Once the nucleus is formed, monomers are templated into growing amyloid fibers. This process can be visualized by an increase in ThT fluorescence. When the monomer population is depleted, elongation stops and enters the stationary phase. (LeVine, 1999a)

2.7 Differentiating “Nucleation and Seeding”

Nucleation requires soluble monomers to undergo conformational changes to form a potent nucleus having the propensity to sequester other monomeric soluble proteins translating into a repetitive oligomer and form larger polymers. The process of nucleation can occur randomly in the case of disease related amyloid for example in the case of Alzheimer´s disease (Harper and Lansbury, 1997; Jarrett and Lansbury, 1993; Naiki and Gejyo, 1999) or can be taken over by the specific nucleator protein as in the case of curli, where a dedicated nucleator protein exists conversion of monomers can also be initiated by preformed amyloid fibrils to induce faster aggregate formation this process is termed seeding (Agrawal et al., 2011; Hammer et al., 2007a; Harper and Lansbury, 1997; Jarrett and Lansbury, 1993; Naiki and Gejyo, 1999). (**Figure 5**)

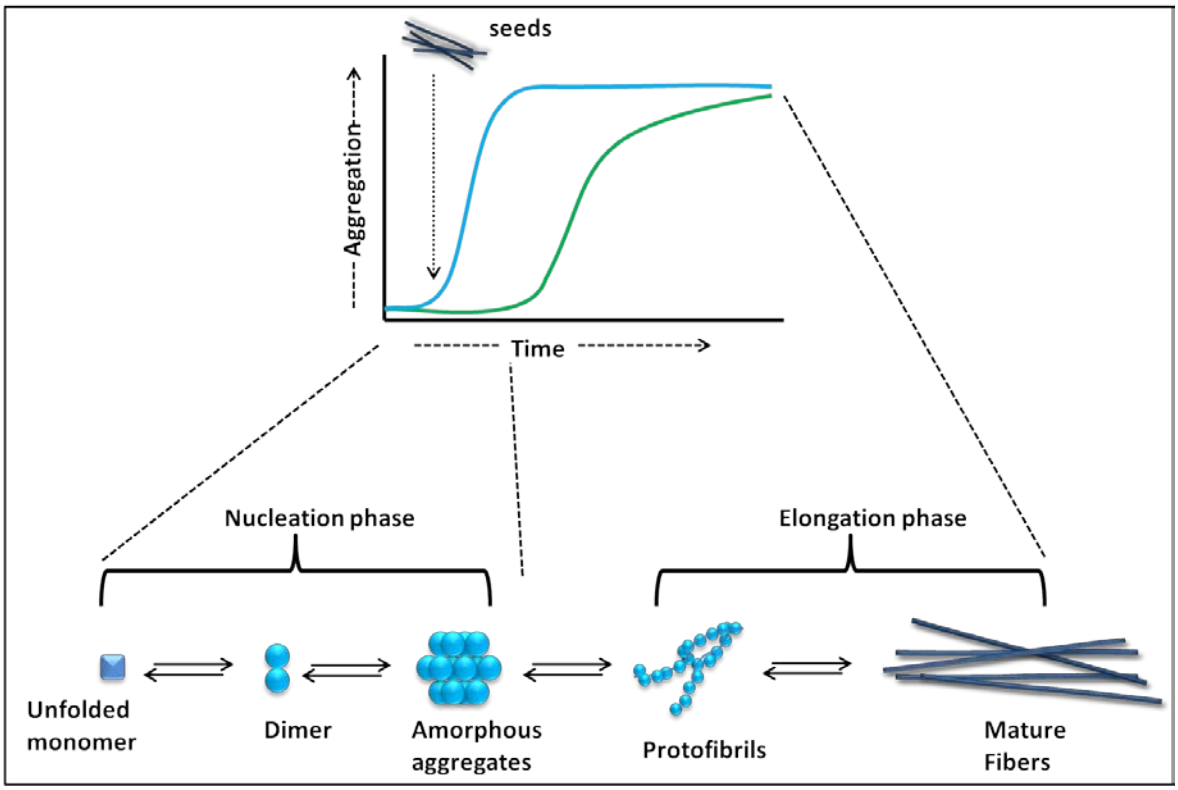


Figure 5: Nucleation-dependent polymerization model of amyloid aggregation.

The kinetics of amyloid formation is well represented by a sigmoidal curve with a lag phase followed by rapid growth phase **(green curve)**. The rate limiting step in the process is the formation of nuclei/seeds to promote aggregation. Hence, addition of preformed seeds significantly increases the rate of amyloid formation. The addition of seeds reduces the lag time and induces faster aggregate formation **(blue curve)**.

2.8 Amyloid toxicity

Discovery of Pike et al., (Pike et al., 1991), established that harmless monomers of Aβ become neurotoxic upon aggregation. The toxicity of Aβ due to polymerization of monomers in to oligomers and higher aggregated forms has been previously reported (Lorenzo and Yankner, 1994). This is further supported by *in vitro* (Hartley et al., 1999; Lambert et al., 1998; Townsend et al., 2006) , and *in vivo* studies showing that oligomeric and pre-fibrillar Aβ assemblies are potent neurotoxins causing synaptic loss and severity of cognitive impairment. Thus, the characteristic neurotoxic pathology is a result of accumulation of toxic oligomeric assemblies of Aβ which are not cleared from the brain.

2.9 Amyloid Folds

Amyloids show different structural folds as observed by fiber diffraction (Jahn et al., 2010; Sunde et al., 1997a) , X-ray crystallography (Nelson et al., 2005; Sawaya et al., 2007) and fluorescence labeling studies. Amyloid folds can be classified in to three broad categories. Example- Antiparallel, parallel in-register and β-solenoid arrangement. In amyloid fibrils the parallel organization is most common (Shewmaker et al., 2009; Toyama and Weissman, 2011a) and the β-strands are coordinated in register or in a β-solenoid.

Parallel in register beta sheets:

These fibrils constitute stacked polypeptide strands that are oriented perpendicular to the fibril axis and form backbone hydrogen bonds with the adjacent strands aligned in-register and parallel. Each residue forms an aligned row along the long axis of the fibrils. As a result, the main-chain hydrogen bonds run parallel to the long axis and β sheets arrange along the length of the fibril (**Figure 6 a**).

β Solenoids / β Helices:

The strands in a β helix (or solenoid) align to form parallel β sheets, but the strands wrap around an axis in a helical arrangement. Two types of β-solenoid include the β-helix (**Figure 6c**), where three β-sheets form a triangular interface, and the β-roll (**Figure 6 c**), where two sheets form an interface akin to the β-sandwich. An example of a β-solenoid structure is represented by the [Het-s] prion.

Dry 'steric zipper':

A beta sheet motif was postulated for the first time by Eisenberg in his recent work (Sawaya et al., 2007) which revealed atomic-level structures of the cross-β spine, a fundamental unit of amyloid-like fibrils. The steric zipper is formed by two tightly interdigitated β-sheets consisting of fibril-forming two of many fibril forming segments G**N**N**Q**Q**N**Y and N**N**Q**Q**NY from the yeast prion protein Sup35 (Nelson et al., 2005). Side chains Asn2, Gln4 and Asn6 point inwards (bold), forming the dry interface. Unlike normal protein, the interface between the two sheets is dry due to the absence of water molecules, with the wet interfaces on the outside surfaces. This form of beta fold may be associated with robustness and self nucleating capacity of diseased amyloid.

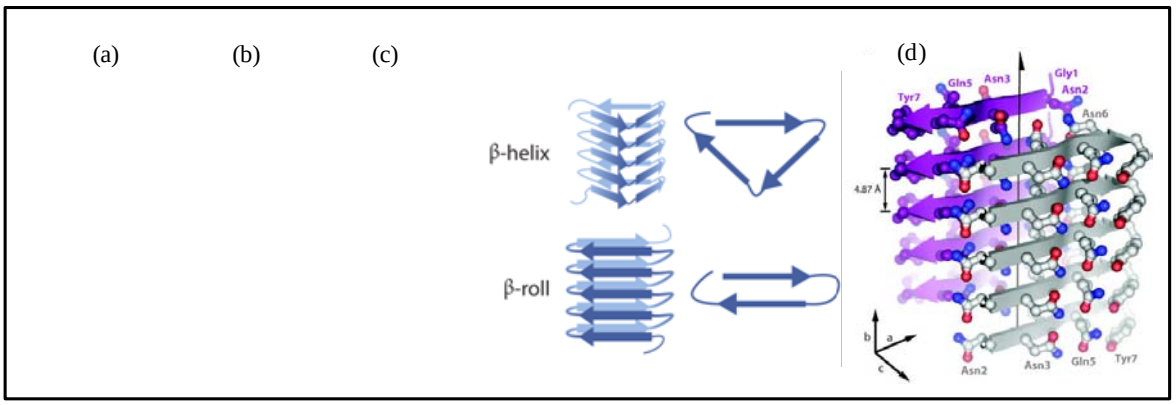


Figure 6: Amyliod folds

(a) antiparallel beta sheet **(b)** parallel in-register beta sheet **(c)** Types of β-solenoids are depicted schematically include the β-helix and β-roll (adapted from(Toyama and Weissman, 2011a). **(d)** Atomistic structure of microcrystals of the peptide GNNQQNY indicating the "dry steric zipper"(Nelson et al., 2005)

3. Curli biogenesis- Division of Labour

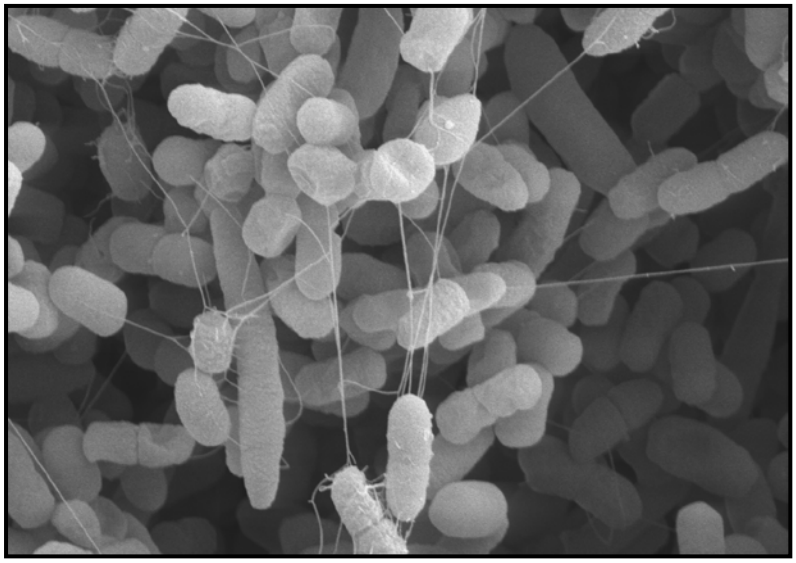


Figure 7: Scanning electron microscopy of CURLI bacterial biofilm of E.coli (UTI strain). The long fibers shown maintain cell to cell interaction. (**Taken by Manfred Rohde).**

Of all the functional amyloid systems known, one of best explained and most intriguing system is curli. Curli constitutes the major proteinaceous component of a complex extracellular matrix produced by many *Enterobacteriaceae* such as *Escherichia coli* and *Salmonella* spp (Barnhart and Chapman, 2006a).

Curli were first discovered in the late 1980s on Escherichia coli strains implicated in bovine mastitis, and have been involved in many physiological and pathogenic processes (Olsen et al., 1989). Curli fibers mediate many cellular processes like biofilm formation which convey resistance to desiccation, and promote long-term survival (Barnhart and Chapman, 2006b; Gerstel and Romling, 2003; Uhlich et al., 2006). Curli is also involved in host cell adhesion, invasion, and immune system activation (Bian et al., 2000; Bian et al., 2001; Gophna et al., 2001; Kikuchi et al., 2005; Uhlich et al., 2006; Vidal et al., 1998; Wang et al., 2006). Curliated bacteria have been shown to colonize chemically diverse surfaces such as plant tissues, stainless steel and glass (Barak et al., 2005; Jeter and Matthysse, 2005; Ryu et al., 2004) (**Figure 7**).

In contrast to the aberrant protein folding giving rise to disease-associated amyloid, the production of the functional amyloid curli is the product of a highly regulated and directed process. Here, different proteins are involved in fiber growth (CsgA), fiber nucleation (CsgB), transcriptional activation (CsgD), outer membrane pore (CsgG) for protein

secretion to the extracellular surface, chaperones for keeping the amylidogenic protein unfolded inside the cell (CsgF, CsgE). The complex process of controlled amyloid formation is thus simplified by the division of labour. The fundamental similarities between curli and disease-associated amyloid such as general morphology and certain biophysical properties provide an opportunity to use curli as a template to study disease-associated amyloidogenesis and cytotoxicity.

3.1 **Curli specific proteins**

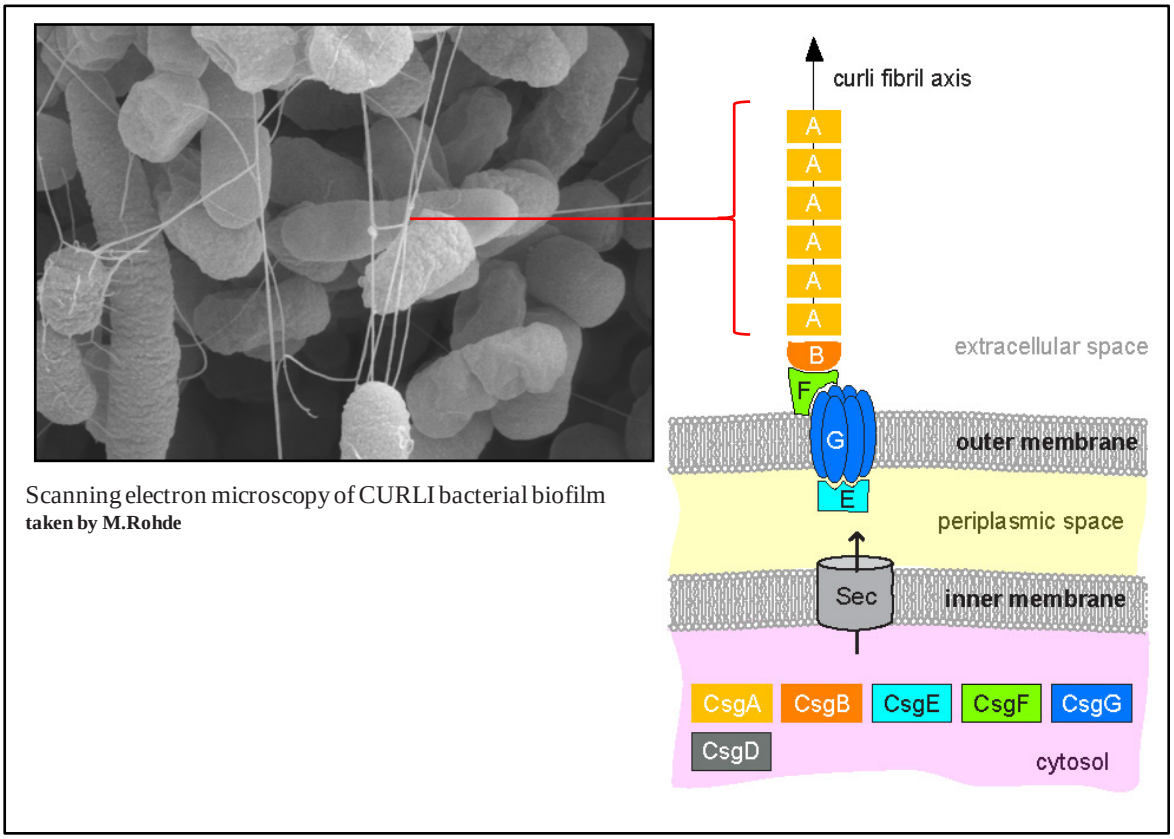
The curli operon codes for major curlin protein CsgA and the minor nulceator protein CsgB (Bian and Normark, 1997; Hammar et al., 1996) incorporating both subunits into the fiber (White et al., 2001) Curli polymerization is thought to occur after secretion of the subunits to the extracellular space, as CsgA and CsgB do not have to be expressed from the same cell in order to assemble curli. In a process called 'interbacterial complementation', CsgA secreted from a donor cell can be assembled into a fiber on the surface of a CsgB-producing acceptor cell (Barnhart and Chapman, 2006b; Hammar et al., 1996). The polymerized, curli fibers exhibit the biochemical and structural properties of amyloids (**General Amyloid features**). Curli is assembled by a unique and highly regulated extracellular nucleation/precipitation pathway. Curli formation in E. coli involves at least six proteins, encoded by the divergently transcribed csgBA and csgDEFG operons (csg, curli specific genes) (Barnhart and Chapman, 2006b; Hammar et al., 1995b) (**Figure 8**).
Along with the csgBA operon a third gene, csgC has been shown, but no transcript for csgC has been detected and no reported role of it has been described in curli biogenesis (Collinson et al., 1996; Hammar et al., 1995a). However mutant strains lacking CsgC still assemble curli but they show defects in auto-aggregation, form paler colonies on Congo red agar plates, and display variable affinity for soluble fibronectin (Barnhart and Chapman, 2006a; Hammar et al., 1995a). Other mutation studies have indicated that AgfC a homologue Tafi protein in *Salmonella Enteritidis* was localized in the periplasm and that it is important for AgfA extracellular assembly, facilitating Tafi synthesis (Gibson et al., 2007).

CsgD is a positive transcriptional regulator in the FixJ/UhpA family. Expression of csgBA absolutely requires CsgD (Barnhart and Chapman, 2006b). The regulation of curli gene expression is complex and is influenced by many environmental cues. For laboratory

strains of *E. coli* and most *Salmonella* strains, curli expression is best at temperatures below 30 °C (Arnqvist et al., 1992) Interestingly, many clinical strains of *E. coli* including sepsis isolates, can express curli at 37 °C and some mutations in the csgD promoter result in strains that produce curli regardless of temperature (Romling et al., 1998; Uhlich et al., 2001). In addition, low salt condition and nutrient limitation (nitrogen, phosphate and iron) stimulate curli gene expression (Gerstel and Romling, 2001; Romling et al., 1998). Oxygen tension also plays a role in curli expression, with microaerophilic conditions resulting in maximal csgD transcription (Gerstel and Romling, 2001). There are many transcriptional regulators modulating the csgDEFG transcription.OmpR/EnvZ, RpoS, Crl, MlrA and IHF positively activate the transcription of csgDEFG. CpxA/R and Rcs negatively influence the transcription of csgDEFG (Barnhart et al., 2006).

CsgE, CsgF, CsgG, CsgC, CsgA and CsgB contain an N-terminal Sec-signal sequence which directs these proteins into the periplasmic space through the SecYEG machinery (**Figure 8**). CsgG is an outer membrane lipoprotein and is proposed to be the major component of the curli-specific secretion machinery. In the absence of CsgG, curli fiber subunits CsgA and CsgB are not stable (Loferer et al., 1997). Expression of both CsgG and CsgA resulted in secretion of CsgA into the extracellular milieu (Barnhart and Chapman, 2006a). CsgG interacts with CsgE and CsgF at the outer membrane (Robinson et al., 2006). Purified CsgG forms barrel structured oligomers and over expression of CsgG increases outer membrane permeability (Robinson et al., 2006).

CsgE and CsgF are indispensable for curli assembly (Barnhart and Chapman, 2006a) but their exact functions are only beginning to be elucidated. In the absence of CsgE, the protein levels of CsgA and CsgB are significantly decreased and curli formation is affected (Barnhart and Chapman, 2006a). A CsgF mutant is defective in the nucleation by CsgB and produced less fibers which were found not associated with cells (Barnhart and Chapman, 2006a; Hammer et al., 2007b).



Scanning electron microscopy of CURLI bacterial biofilm
taken by M.Rohde

Figure 8: Schematic diagram of curli gene.

CsgD is a positive transcriptional regulator of the csgBA operon. All the proteins encoded by the csg operons, except for CsgD, contain sec signal sequences for translocation into the periplasm. Outer membrane protein CsgG required for the secretion of the two curli structural subunits CsgA and CsgB. CsgA is secreted outside of the cell where CsgB nucleates it into a fiber. CsgE and CsgF both interact with CsgG and are required for efficient curli assembly.

3.2 Curli Structure

The primary structure of the subunits of CsgA and CsgB can be divided into three parts, a Sec signal sequence, an N-terminal 22 or 23 residues and five imperfect repeating units (R1, R2, R3,R4 and R5) (**Figure 9**). The Sec signal sequences are cleaved when they pass through the inner membrane (White et al., 2001) (Collinson et al., 1999b). Therefore, the mature CsgA and CsgB only contain an N-terminal flexible region and five repeating units. The N-terminal stretch is sensitive to proteinase K treatment in CsgA, whereas the five repeating units are resistant indicating that they comprise the amyloid cores (White et al., 2001) (Collinson et al., 1999b). The N-terminal stretch of CsgA was reported to target CsgA to the CsgG secretion machinery while the role of the N-terminal stretch of CsgB remains elusive (Robinson et al., 2006).

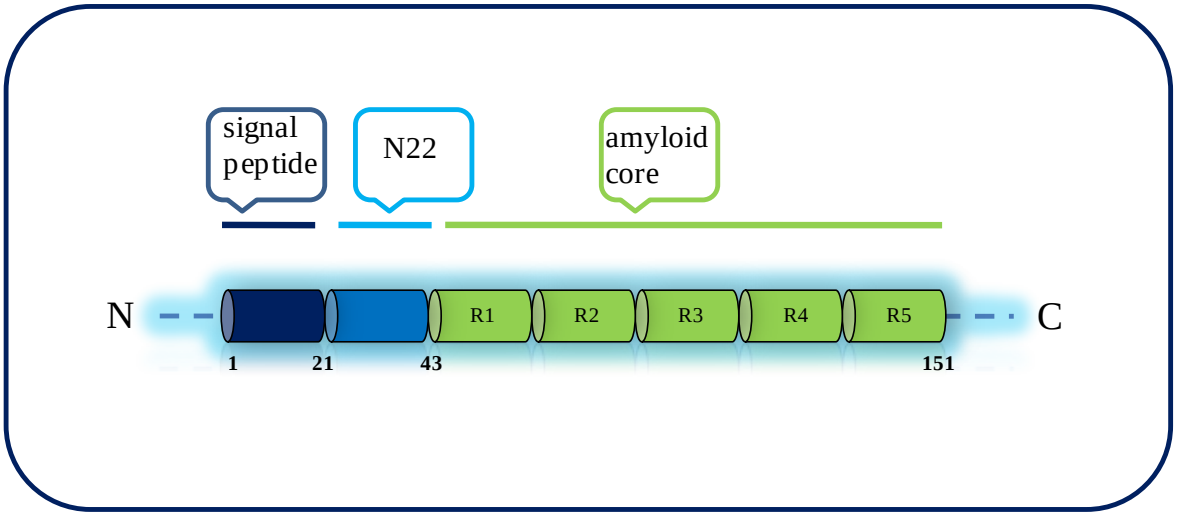


Figure 9: Structural domain of curli protein CsgB & CsgA.

Mature protein consists of a flexible N-terminus and an amyloid core made of five repeats with a characteristic consensus sequence.

The repeating units of CsgA contain a consensus sequence (S-X5-Q-X-G-X2-N-XA-X3-Q). Each repeating unit is predicted to form a β strand-loop-β strand motif, resulting in five repeats forming a cross-β structure (**Figure 10**) (Collinson et al., 1999b). Similar to CsgA, the first four repeating units of CsgB have a consensus sequence (X6-Q-X-G-X2-N-X-A-X3-Q) and the five repeating units are predicted to form a cross-β structure (White et al., 2001). The R5 of CsgB is different from other repeats in that it does not have conserved Gln and Gly residues (**Figure 10**). Although CsgA and CsgB share similar amino acid sequences and predicted structural features, the theoretic isoelectric points (pI) of mature CsgA and CsgB are 4.51 and 9.30, respectively, because CsgB has significantly more cationic residues such as Arg and Lys than CsgA, especially in R5 (**Figure 10**). Unlike CsgA, CsgB is associated to the outer membrane and is mostly tethered to the cell surface (Hammar et al., 1996) (Loferer et al., 1997)

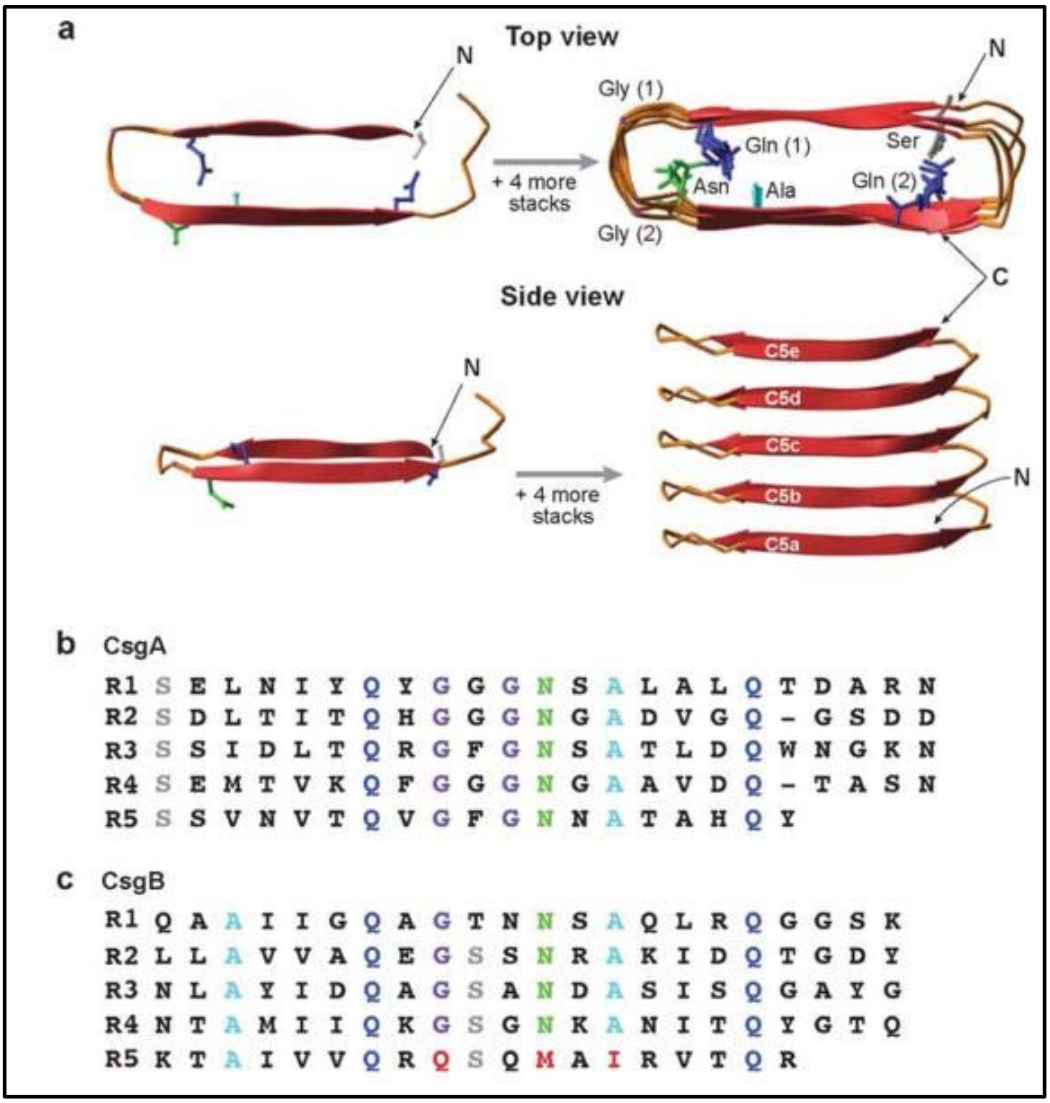


Figure 10: Model of CsgA and CsgB structure.

(adapted from (Barnhart and Chapman, 2006a)

(a) A model of the predicted strand-loop-strand motif of CsgA (Collinson et al., 1999a) . A similar prediction was made for CsgB (White et al., 2001) **(b)** CsgA and **(c)** CsgB are composed of five repeating units (R1–R5), and each repeating unit is the equivalent of one strand-loop-strand. Residues that are conserved within the repeating units are colored similarly. The red residues in R5 of CsgB indicate residues that differ from R1–R4 of CsgB.

The deletion of the last repeat of CsgB resulted in a molecule disassociated from cells (Hammer et al., 2007b). Lipopolysaccharides (LPS) are the major component of the outer membrane of Gram-negative bacteria.

3.3 Molecular mechanism of Curli polymerization

Like disease-associated amyloids, curli are non-branching, beta-sheet rich fibers that are resistant to protease digestion and denaturation by 1% SDS (Barnhart and Chapman, 2006a; Collinson et al., 1993; Collinson et al., 1991; Collinson et al., 1999b; Olsen et al., 1989). Amyloid fibers, including curli, are propagated by the addition of soluble precursors to the growing fiber tip in a process called elongation (Agrawal et al., 2011; Naiki et al., 1998; Serio et al., 2000). The polymerization of purified CsgA in vitro displays a lag phase of 1-2 hour that precedes rapid fiber elongation (LeVine, 1993). Addition of preformed CsgA fibers significantly shortens the lag phase (Agrawal et al., 2011). The CsgA fiber seeds added provides specific template that guides fiber elongation and allows polymerization to proceed with a shorter lag phase. The amyloid seeding is usually occurs between proteins and fibers with identical or nearly identical amino acid sequences (O'Nuallain et al., 2004).

Both CsgB and CsgA contain a conserved five-fold sequence homology within their amino acid sequences (Barnhart and Chapman, 2006b; Collinson et al., 1999b; Hammar et al., 1996; White et al., 2001). The regions of homology called repeating units in CsgA have recently been demonstrated to be amyloidogenic (Agrawal et al., 2011).
Although CsgA polymerization can be initiated by fiber seeds in vitro, in the in vivo conditions, assembly of CsgA polymers on the bacterial cell surface is dependent on the amyloidogenic CsgB nucleator protein (Hammer et al., 2007b). Amino acid sequence of CsgA and CsgB shows 30% identity and in silico molecular modeling predict a similar cross beta-strand structure (Collinson et al., 1999b; White et al., 2001) .

4. Techniques to Study Amyloid Structures

Understanding the protein function at the molecular level would require high resolution structural characterization of the protein. Due to the large size of the fibril aggregates, dynamic conformational exchange during the aggregation process and intrinsic heterogeneity of the high molecular weight assemblies and their precursors, studies of protein aggregation present a great challenge to the structural biologist. X-ray crystallography struggles with the insolubility, non crystalline nature and complexity of the

fibrils. Cryo electron microscopy falls short in providing images at atomic resolution; however nuclear magnetic resonance (NMR) proves to be the method of choice in investigation of amyloids. There are other low resolution techniques which would provide information on the secondary structure of the amyloids. Some of the low and high resolution techniques to study the structure of the amyloid fibrils will be discussed below. Recent research has adapted known techniques for structure analysis in order to overcome the challenging properties of amyloids (Chiti and Dobson, 2006a; Toyama and Weissman, 2011b), as for example hydrogen/deuterium (H/D) exchange NMR.

4.1 Low Resolution Techniques

Techniques like atomic force microscopy and transmission electron microscopy display the fibrous component consisting of a number of filaments side by side parallel to each other longitudinally while the X-ray fiber diffraction shows the distance measurement within the beta sheets of the fibers (Harper et al., 1997; Sunde et al., 1997b). Fourier transform infrared (FT-IR) and circular dichroism (CD) spectra help to verify a β-sheet rich fold by specific absorption bands, as well as the detection of the folding kinetics (Berthomieu and Hienerwadel, 2009; Ranjbar and Gill, 2009). FT-IR spectra identify amyloid specific β-sheets by a characteristic wave number of the amide I band at 1625 – 1610 cm-1, compared to soluble proteins with a wave number ≥ 1630 cm-1 for beta sheets (Zandomeneghi et al., 2004). Beta-sheet shows a characteristic CD spectrum consisting of a negative band at 218 nm and a positive band at 195 nm. Amyloid specific dyes like Congo red and Thioflavin T binds to various sub-populations of the folding intermediates and fibers (Hawe et al., 2008), which can be used to monitor aggregation kinetics, surface hydrophobicity (Hawe et al., 2008) and tertiary and quaternary structure of the amyloid fold.

Other indirect techniques include proteolytic digestion where the amyloid core resists digestion by the proteolytic enzyme and remains protected unlike the flexible region of the protein (Kheterpal et al., 2001). Several approaches make use of mutagenesis which includes introduction of proline as a β-breaker (Chiba et al., 2003; Williams et al., 2004; Wood et al., 1995) and their influence on the overall fold helps in identifying the stable regions of the amyloid core. Introduction of single cysteine residues in the protein for electron paramagnetic resonance (EPR) or and fluorescence measurements (118-120) by using thiol specific paramagnetic spin labels or extrinsic fluorescence dyes can reveal the

amyloid core (Chen et al., 2007; Der-Sarkissian et al., 2003; Torok et al., 2002). The above methods would provide information about the presence or absence of regular secondary structure elements at a defined residue, the topology of β-sheets, as well as inter- and intramolecular contacts within the fibril.

4.2 High Resolution Techniques

Recent advances in various biophysical methods facilitate a closer insight into amyloid structures by cryo electron microscopy, X-ray crystallography, as well as NMR spectroscopy techniques such as hydrogen/deuterium (H/D) exchange NMR or solid state NMR.

In cryo electron microscopy (cryo EM) the samples are vitrified with liquid nitrogen and measured at the cryogenic temperatures enabling the analysis of specimens in their native environment without staining or fixing. In Cryo-Electron Microscopy, a stream of high energy particles bombards the sample. The image produced is a result of the interaction of the sample with this electron beam (Henderson, 2004). The observed electron densities reveal the helical twist of the protofilaments arranged side by side and the cross-β structure of amyloids (Jimenez et al., 1999; Jimenez et al., 2002; Meinhardt et al., 2009; Sachse et al., 2008; Schmidt et al., 2009; Zhang et al., 2009).

Even though fibrils formed by amyloidogenic peptides and proteins are highly ordered structures, yet their degree of order is not high enough to produce high resolution diffraction patterns as the way single crystals do.

Eisenberg and co-workers were able to use X-ray crystallography to analyse microcrystals of peptide fragments forming amyloid-like structures and revealed atomistic details congruent to existing structural models of amyloids (Nelson et al., 2005; Sawaya et al., 2007; Wiltzius et al., 2009; Wiltzius et al., 2008). Thus, the β-sheet stacking interface “dry steric zipper” was postulated by Eisenberg as a basic principle of the cross-β structure (Nelson et al., 2005; Sawaya et al., 2007). However, these studies are limited to short peptides of seven residues or less, as longer peptides or even proteins do not crystallize into an amyloid-like packing.

4.3 Theoretical Background of NMR Spectroscopy

NMR has proved to be the best way for determination of three-dimensional (3D) structures of non-crystallisable proteins and protein complexes at atomic resolution. NMR spectroscopy is widely used for the structure elucidation of small chemical compounds. It can also be applied for structure determination and structure-function relationship analysis of proteins under physiological conditions (Wüthrich, 1990). In addition to conventional structure determination of proteins, NMR enables the investigation of folding intermediates (Hofmann et al., 2008; Schulenburg et al., 2009), chemical properties of functional groups (Gardiennet et al., 2005; Shammas et al., 2007), intermolecular interactions (Debnath et al., 2011; Drechsler et al., 2011), hydrogen bonding (Assadi-Porter et al., 2003; Liwang and Bax, 1997) as well as structural investigation of amyloids and membrane proteins (Hoshino et al., 2002; Linser et al., 2011; Tycko, 2011).

Conventional structural elucidation using solution NMR is limited by the size of the protein (~30kDa) and its solubility for standard applications. However, specific application within the field of NMR spectroscopy exits that are applicable to amyloid fibrils making this the only technique to date that can provide structural information at atomic resolution on amyloids formed by full length protein.

The principle behind NMR is that many nuclei with an odd mass or odd atomic number have "nuclear spin" in multiples of 1/2 and can be + or -. This includes ^{1}H and ^{13}C (but not ^{12}C). If an external magnetic field is applied, an energy transfer is possible between the base energy (+1/2) to a higher energy level (-1/2). The energy transfer occurs at a wavelength that corresponds to radio frequencies and when the spin returns from the excited state to its base level, energy is emitted at the same frequency. The signal that matches this transfer is measured and processed in order to yield an NMR spectrum for the nucleus concerned (Hornak, 1997-2011).

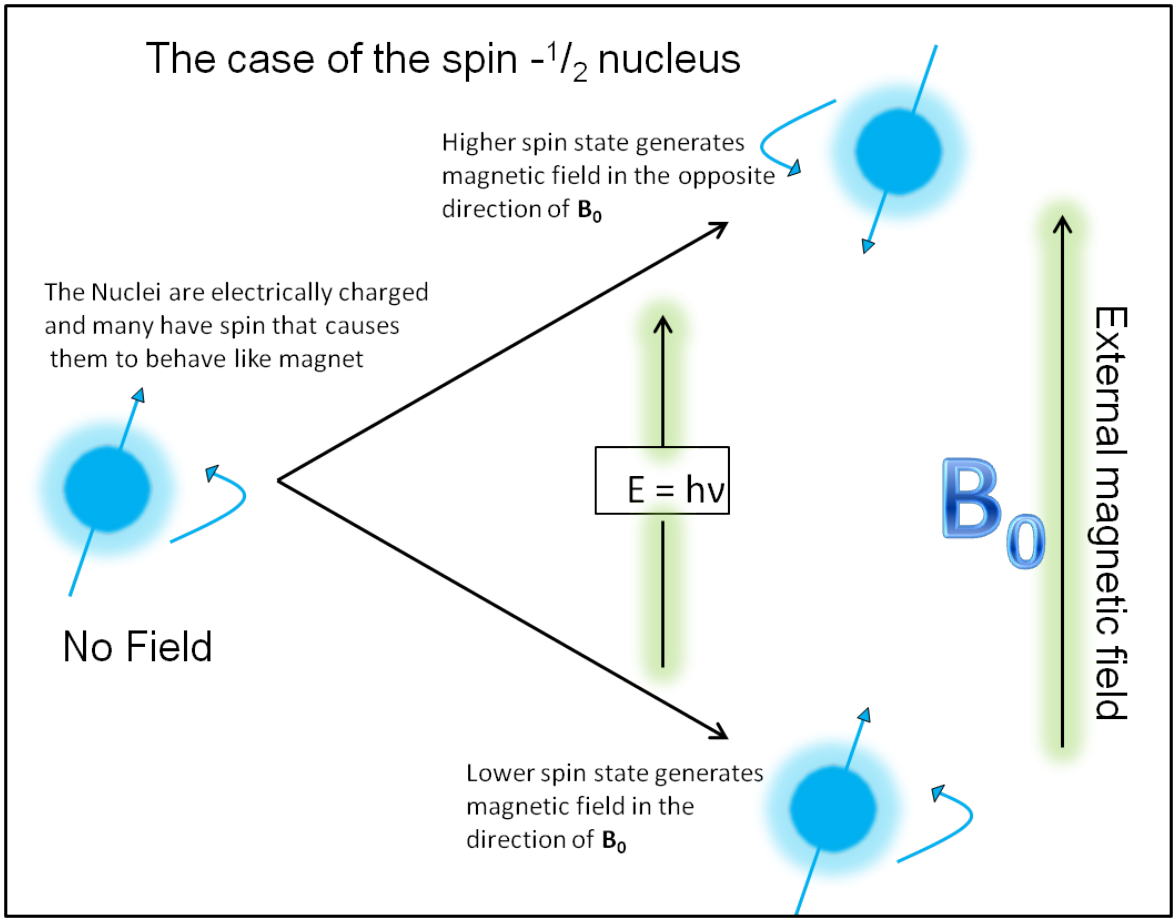


Figure 11: NMR energy level diagrams of nuclear spin states.

The spin-½ nuclei that include the most commonly used NMR nucleus, proton (1H or hydrogen-1) as well as many other nuclei such as 13C, 15N and 31P.

When placed in a magnetic field of strength B, 1/2 nuclei with a net spin can absorb a photon, of frequency ν. The frequency ν depends on the gyro-magnetic ratio, γ of the 1/2 nuclei. (**Figure 11**)

$$\nu = \gamma B$$

The energy, E, of a photon is related to its frequency, ν, by Planck's constant (h = 6.626x10-34 J s). Where ν = γ B and E = h ν, therefore the energy of the photon needed to cause a transition between the two spin states is equation (1)

$$\mathbf{E = h \gamma B} \qquad \text{(Equation 1)}$$

When the energy of the photon matches the energy difference between the two spin states absorption of energy occurs following the saturation of the higher energy level and relaxation when the spin returns to its base level, energy is emitted at the same frequency. The energy difference between the two spin states also depends on its chemical environment which can influence the applied magnetic field B on a nucleus. The

neighbouring nuclei and their electrons act as tiny magnets and can either shield or deshield the applied magnetic field on the nucleus causing variation of the resonance frequency. Different nuclei can be distinguished by their resonance frequencies, which are influenced by the chemical environment (Hornak, 1997-2011).

Important nuclei in NMR spectroscopy are nuclei with spin=1/2 ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P. The natural abundance of ^{13}C and ^{15}N nuclei is relative of low compared to ^{1}H isotopes. For this reason, proteins chosen for NMR experiments have to be labelled with these isotopes.

The NMR technique has been further developed to multi-dimensional (2D, 3D, 4D up to 7D) NMR experiments to overcome the overlap of signals in 1D NMR spectra. These NMR experiments are necessary to solve the complex structure of proteins.

With increasing particle signals originating from protons broaden beyond detection and therefore are not directly observable in the solution-state NMR spectroscopy. This is caused by increased relaxation rates due to contributions from dipolar couplings and chemical shift anisotropy (CSA).

To overcome this problem in solid state NMR the sample is being spun at an angle $\theta m \sim 54.7^{\circ}$, also called the “magic angle”. Under magic angle spinning (MAS), CSA and dipolar couplings are removed (if the sample is spinning “fast enough”) and spectra with relatively narrow lines can be recorded. However structure elucidation of protein by solid state NMR is technically still very challenging and the first structure of a small model protein has been solved merely ten years ago.

4.4 Quenched H/D exchange

Due to the limitation of the study of the fibrils by solution state NMR methods like quenched H/D exchange [^{15}N, ^{1}H] NMR spectroscopy in combination with a sequence specific assignment of resonances is a powerful tool to characterize structured regions in proteins at single residue resolution (Englander, 2000; Li and Woodward, 1999).

The HDX method uses standard solution-state NMR. This method is based on the principle that a protein placed in a deuterated solvent results in an exchange of the amide protons from the protein with the deuterium in the surrounding. This proton to deuterium exchange is based on the position of the protons on the protein. The amide protons involved in the secondary structure display a slow exchange compared to the other amide protons in the protein backbone (Ippel et al., 2002) .

In this case, when [^{15}N, ^{1}H] labelled amyloid fibers are incubated with the deuterated buffer there is an exchange of the amide protons of the protein with the deuterium of the surrounding solvent (Englander et al., 1997; Hvidt and Linderstrøm-Lang, 1954). This exchange is influenced by the chemical environment of the amide proton in the protein (Bai et al., 1993). For example, backbone amide protons, which are involved in the formation of hydrogen bonds in the secondary structure of the protein, are less accessible to the solvent while the free amide protons are easily exchanged with the solvent. Amino acid residues where the amide protons have exchanged with deuterium become NMR invisible (Hornak, 1997-2011). Hence, recording a 2D NMR spectrum correlating the proton and nitrogen (HSQC) provides information about the protonation of backbone amides and their exchange behaviour. This results in the identification of the protein amide backbone involved in the formation of the secondary structure, allowing distinction between ordered and disordered regions of an amyloid.

The H/D exchange NMR experiment of amyloids includes three distinct steps (Hoshino et al., 2002):

A) Incubation of the amyloid fibrils in D_2O buffer
B) Monomerisation of the fibrils with DMSO
C) NMR analysis

First, the amyloid fibrils are incubated in D_2O buffer in order to facilitate Hydrogen/Deuterium exchange between the protein and the solvent. Subsequently, the reaction is quenched at specific time points by removing the solvent and freezing the fibrillar proteins in liquid nitrogen. Secondly, the fibrils are monomerised by using a mixture of deuterated DMSO + 0.1 % tri-fluroacetic acid. Importantly, the aprotic DMSO conserves the hydrogen/deuterium pattern present in the fibril at the selected time point. Finally, [^{15}N, ^{1}H] correlation spectra (HSQC) are recorded and analysed for residual intensity of resonances in comparison to spectra of the non exchanged fibrils. The backbone amides corresponding to the absent resonances have undergone a fast exchange as deuterium is not detectable in such NMR spectroscopy.

Furthermore the H/D exchange NMR technique has been improved in different ways (Luhrs et al., 2005; Ritter et al., 2005). One improvement is the measurement of series of [15N, 1H] correlation spectra in order to correct the resonance intensities for exchange in DMSO. Remaining traces of D_2O are able to mask the resonance intensities due to further exchange in DMSO. For this reason, a rapid mixing with DMSO and pre-adjusted NMR

settings (together less than 1 min) are of great importance. The acquisition of one spectrum was reduced to less than 5 min. Another aspect is the analysis of NMR spectra after different incubation intervals in D_2O. These data enable the accurate examination of exchange rates of each assigned backbone amide. Importantly, the exchange rates provide additional information about the homogeneity and regularity within the fibrils (Luhrs et al., 2005; Ritter et al., 2005).

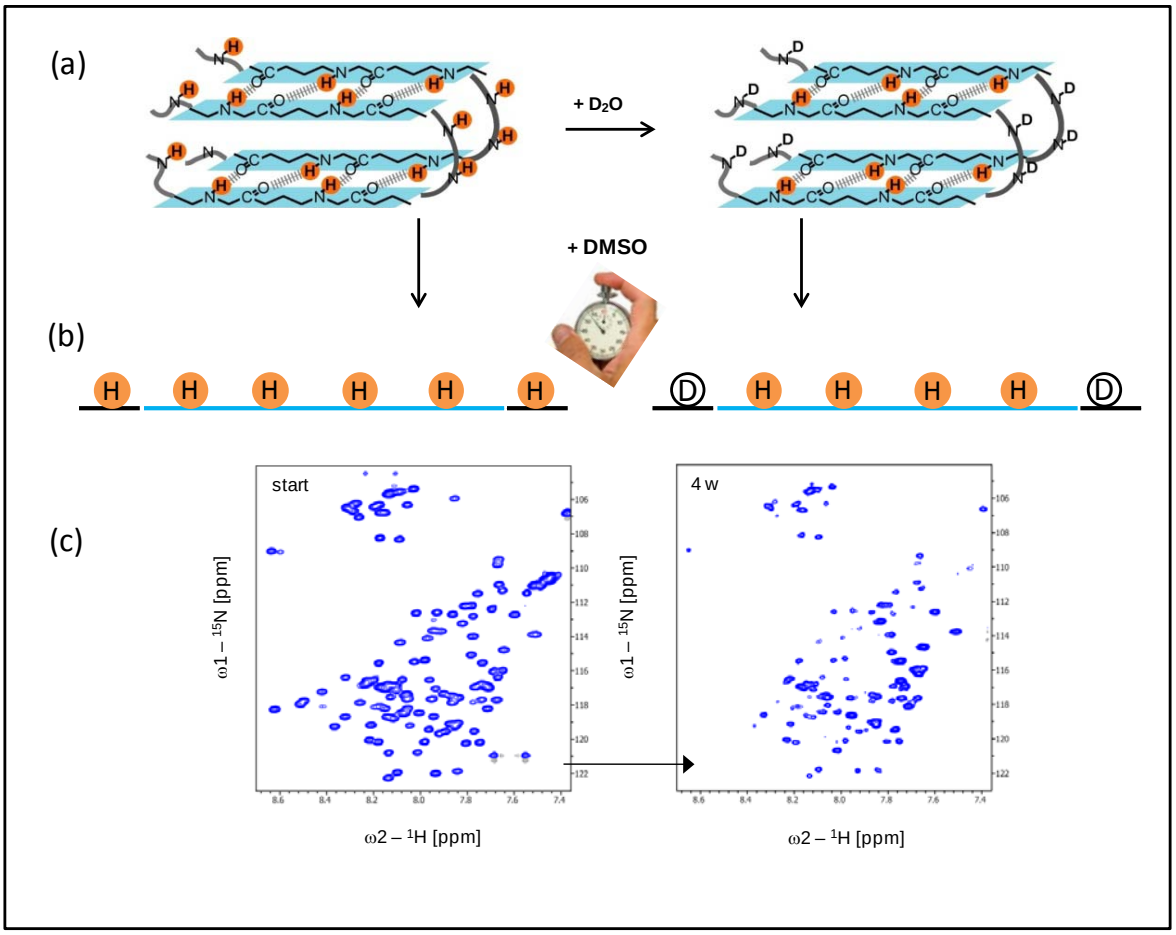


Figure 12: Schematic description of the H/D exchange NMR experiment.

(Adapted from Christiane Ritter)

(a) During the H/D exchange period fully protonated fibrils incubated with D_2O. Hydrogen bonded protons are protected form exchange, while the protons in the flexible region exchange to D_2O. **(b)** Monomerisation and unfolding of the amyloid fibrils using DMSO. Original positions of β-strands are highlighted in the same colour as in **(a)**. **(c)** Detection of H/D exchange via NMR recording [15N, 1H] correlation spectra. Obtained NMR spectra differ in their peak intensities.

In summary, quenched hydrogen/deuterium exchange (HDX) NMR spectroscopy is a unique tool to study stability and thermodynamic behaviour of proteins and their variants. This approach was already successfully applied to a number of systems to map the amyloid cores of HET-s (Ritter et al., 2005), Sup35 (Toyama et al., 2007), Aβ (Luhrs et al., 2005),

β_2-microglobulin (Hoshino et al., 2002), α-synuclein (Vilar et al., 2008) and SH3 domain fibrils (Carulla et al., 2005).

The robustness of HDX experiments have been verified by mass spectrometry HD-MS (Kheterpal et al., 2006). In this method after the HDX, the fibrils are analysed my mass spectrometry instead of solution NMR since the higher mass of deuterium enables the detection of the exchange by mass spectrometry. The fibrils after HDX are quenched followed by the digestion of the protein into smaller peptides. Subsequently the digested peptides are separated and analysed by mass spectrometry (Kheterpal et al., 2006). One advantage of HD-MS experiments is the requirement of less protein material without any labelling. Despite the advantage, the major drawback is the process of protease mediated protein digestion in to small peptide fragments suitable for analysis, which is a time consuming process favouring the back exchange of protons. In case of amyloids, solubilisation of amyloids to monomer by strong denaturants like hexafluoroisopropanol, formic acid or DMSO may not favour protease digestion. For the prion forming domain of HET-s from *Podospora anserina* HDX experiments were analysed and compared for both techniques illustrated in **Figure 13** (Nazabal and Schmitter, 2006; Ritter et al., 2005). The comparison indicated obvious differences in the resolution of the techniques.

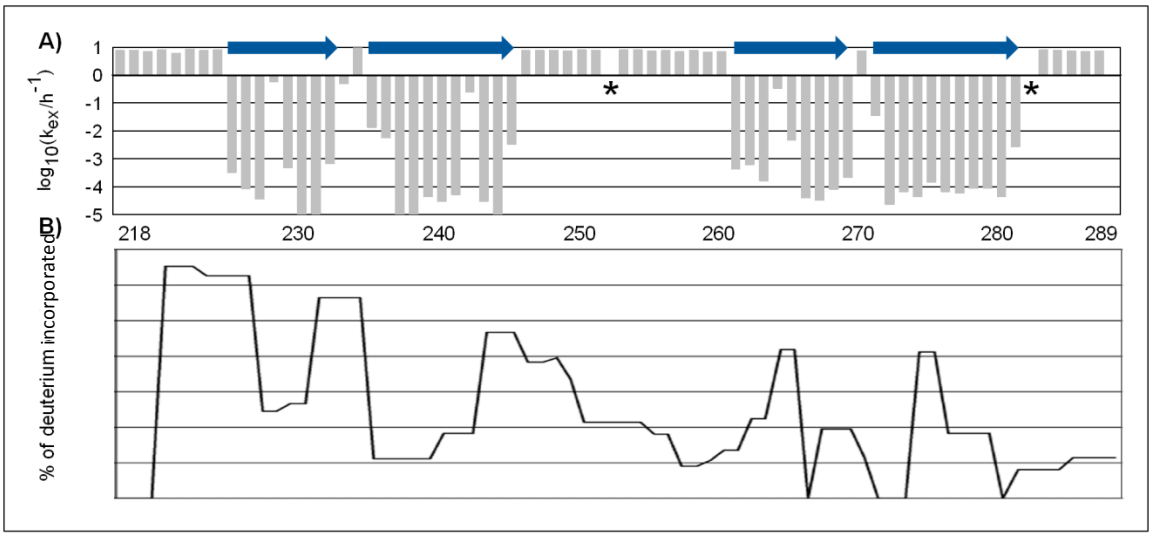


Figure 13: HDX analysis of HET-s218-289.

(a) HDX analysis identified four protected regions assigned to β-strands (Ritter et al., 2005).Bars indicate hydrogen exchange rates k_{ex} (h^{-1}); blue arrows indicate β-strands. Backbone amides involved in hydrogen bonding show slow exchange rates. Black asterisks mark residues, which could not be analyzed. **(b)** MS analysis determined the percentage (%) of deuterium incorporation long the HET-s218-289 sequence (Nazabal and Schmitter, 2006). Less deuterium incorporation indicates protected backbone amide.

5. Aims and Scope

The nature of toxic species in many of the disease related amyloids for example in Aβ has yet to be conclusively identified and little is known about how Aβ polypeptide aggregation begins *in vivo*. Insights into these critical phenomena will undoubtedly lead to advances in AD treatments. Hence exploring systems involving amyloid formation as a natural functional process will provide answers to these questions. Hence, the functional amyloids may hold the key to understanding the molecular mechanisms of amyloid fiber toxicity and initiation of amyloid fiber polymerization.

One of the well studied functional amyloid assembly systems is curli, which are extracellular amyloids produced by several bacterial species, including *Escherichia coli* and *Salmonella spp.* (Barnhart and Chapman, 2006b). Functional amyloid curli mediate biofilm formation, host attachment, invasion and activation of host system immune system. Contrary to the disease related amyloid, these proteins mediate normal cellular functions allowing them to be studied in their native organism. This unique amyloid biogenesis provides an opportunity to use curli as a template to study amyloid toxicity related to neurodegenerative disease. The amyloidogenesis of the major curli subunit protein CsgA is dependent on the minor curli subunit nucleator protein CsgB *in vivo*. Hence the nucleation forms the rate-limiting step in amyloid propagation and its nature remains poorly understood.

The goal of this dissertation was to elucidate the nucleation process in curli. To achieve the following aim, the first hurdle was to standardize the construct coding for the full length curli nucleator protein CsgB and to establish a robust purification procedure to facilitate biochemical and structural studies. Following purification, CsgB was characterized by HDX NMR for its secondary structure. In addition, the structure analysis of CsgB was also assessed for its folding or aggregation kinetics and its role in nucleation of CsgA. CsgB was further probed by a mutagenic approach to create various truncations mutants based on the structure. These mutants were then characterize for their folding and nucleation capacity by using various methods like electron microscopy, fluorescent dye binding to elucidate the role of CsgB in structure-function relationship of curli nucleation.

6. Material and Methods

6.1 Standard Materials

6.1.1 Chemicals, Enzymes, Antibodies and Kits

If not stated otherwise, all chemicals that were used in this work were of *"pro analysis"* grade. Chemicals were purchased from the companies Amersham Biosciences, Fluka, Merck, Millipore, Omnilab, Qiagen, Roche, Roth, Sigma and Stratagene.
Enzymes (**Table 3**), antibodies (**Table 4** and **Table 5**) and kits (**Table 6)** were used as described by the corresponding company.

Table 3: **Enzymes used in this work.**

Enzyme	Source	Specification
BamHI, DpnI, HindIII, NotI, SacI, SalI,SapI, TatI, XbaI	New England Biolabs, Fermentas	Restriction endonucleases
DNAse	Merck	Endodeoxyribonuclease
Proteinase K	EuroClone	Serine peptidase
Platinum *Pfx* Polymerase	Invitrogen	DNA polymerase
Pfu Turbo Polymerase	Stratagene	DNA Polymerase
T4 DNA Ligase	New England Biolabs	DNA Ligase
Shrimp Alkaline Phosphatase	Roche	Phosphate hydrolase
Hen white-egg lysozyme	Fluka	Glycoside hydrolase

Table 4: Primary antibodies used in this work.

Name	AB species	Antigen	Dilution	Source
His	Mouse	Hexa-His	1:1500	Novagen

Table 5: Secondary antibodies used in this work.

Name	AB species	A/R species	Conjugate	Dilution	Source
Anti Mouse IgG (H+L) AP Conjugate	Goat	Mouse	Alkaline phosphatase	1:7500	Promega
Anti Mouse IgG (H+L) HRP Conjugate	Goat	Mouse	Horse radish peroxidase	1:2000	Dianova

Table 6: Kits used in this work.

Kit	Source	Usage
Lumi-Light Western Blotting Substrate	Roche	Detection of HRP conjugated antibodies by ECL
QIAquick® Gel Extraction Kit	Qiagen	DNA extraction from gel slices
QuikChange® Site-Directed Mutagenesis Kit	Qiagen	Selective modification of plasmid DNA by mutation
QIAquick® PCR Purification Kit	Qiagen	Purification of PCR products, removal of nucleotides
QIAprep® Spin Miniprep Kit	Qiagen	Plasmid purification

6.1.2 Molecular Weight Standards

Following molecular weight standards were used (**Table 7**).

Table 7: Molecular weight standards used in this work.

Name	Source	Type
Smart Ladder	Eurogentec	Agarose gel DNA marker
PageRuler™ Prestained Protein ladder	Fermentas	SDS-PAGE protein marker
PageRuler™ Unstained Protein ladder	Fermentas	SDS-PAGE protein marker
Unstained Protein Molecular weight marker	Fermentas	SDS-PAGE protein marker

6.1.3 Bacterial Strains

Following bacterial strains were used (**Table 8**).

Table 8: Bacterial strains used in this work.

E. coli strain	Genotype	Source
BL21 (DE3)	*F⁻ ompT gal dcm lon hsdS$_B$(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Stratagene
DH5α	*F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), λ–*	Invitrogen
T7 Express	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10--Tet^S) endA1 Δ(mcrC-mrr)114::IS10*	NEB
MC4100	*F-, [araD139]B/r, Del(argF-lac)169, lambda-, e14-, flhD5301, Δ(fruK-yeiR)725(fruA25), relA1, rpsL150(strR), rbsR22, Del(fimB-fimE)632(::IS1), deoC1*	HZI

6.1.4 Plasmids

Following recombinant plasmids were used in this work (**Table 9**). Plasmid derivatives of pET-11d, pET-15b and pBar100 carried an ampicillin resistance gene.

Table 9: Recombinant plasmids used.

Plasmid	Features	Source
pET-11d $CsgA_{21-151}$	pET-11d derivative encoding *E. coli* $CsgA_{21-151}$	A. Eberth; HZI, Braunschweig
pET-11d $CsgB_{21-151}$	pET-11d derivative encoding *E. coli* $CsgB_{21-151}$	A. Eberth; HZI, Braunschweig
pET-11d $CsgB_{41-151\ NTRUNC}$	pET-11d derivative encoding *E. coli* $CsgB_{41-151}$	this work
pET-11d $CsgB_{21-131\ \Delta 132-151\ CTRUNC}$	pET-11d derivative encoding *E. coli* $CsgB_{21-131}$	this work
pET-11d $CsgB_{41-131\ NTRUNC\&\ CTRUNC}$	pET-11d derivative encoding *E. coli* $CsgB_{41-131}$	this work
pET-11d Sumo- $CsgB_{22-131}$-Sumo	pET-11d derivative encoding fusion protein Sumo-$CsgB_{22-131}$-Sumo	this work

[1)] Sequences of encoded proteins (excepting variants due to site directed mutagenesis) are given in the Appendix.

6.1.5 Oligonucleotides

Oligonucleotides (**Table 10**) were used as primers for site-directed mutagenesis and PCR. They were purchased from MWG.

Table 10: Oligonucleotides used.

Recognition sites for restriction enzymes are underlined and point mutations/codon exchanges are marked in red. For QuikChange site-directed mutagenesis only forward primers are listed.

Name	Sequence (5´ → 3´)	Restriction site/ Mutation/Tag
	Cloning and sub-cloning	
*csg*B-ΔR5 fwd	GGAGATATACCATGGCAGGTTATGATTTAGC	*Nco*I
*csg*B-ΔR5 rev	CGCGGATCCTTAATGGTGATGGTGATGGTGAGTAC CATACTGTGTAATATTTGC	*Bam*HI
*csg*B-ΔN fwd	CATGCCATGGTCTTCATTTAATCAGGCAGCCATAA TTGG	*Nco*I
*csg*B-ΔN rev	CGCGGATCCTTAGTGGTGGTGGTGGTGGTGACG	*Bam*HI
*csg*B-ΔNΔR5 rev	CATGCCATGGTCTTCATTTAATCAGGCAGCCATAA TTGG	*Nco*I
*csg*B-ΔNΔR5 rev	CGCGGATCCTTAATGGTGATGGTGATGGTGAGTAC CATACTGTGTAATATTTGC	*Bam*HI
	Site-Directed Mutagenesis	
*csg*A·Q49A	CTGAGCTGAACATTTACGCGTACGGTGGCGGTAACT C	Q49A
*csg*A·N144A	CTCAGGTTGGCTTTGGTGCCAACGCGACCGCTCATC AG	N144A

6.2 Media and Buffer

The media and buffer, which were used in this study, are summarised in **Table 11** and **Table 12**. All media were sterilised by autoclaving (121 °C, 2 bar, 20 min, Top7000PST, Sauter) and heat-sensitive additives were sterile filtered (pore width 0.2 µM). Antibiotics were added after media cooling below 50 °C. Depending on plasmid and bacterial strain the following antibiotics were used: ampicillin (100 µg/ml).

Table 11: Media used for bacterial culture.

Medium	Composition
Lysogeny Broth (LB)	1 % (w/v) tryptone, 0.7 % (w/v) NaCl, 0.5 % (w/v) yeast extract, adjusted to pH 7.5 with NaOH
Terrific Broth (TB)	1.2 % (w/v) tryptone, 2.4 % (w/v) yeast extract, 0.4 % (v/v) glycerol, 0.1 x TB salts
TB salts	0.17 M KH_2PO_4, 0.72 M K_2HPO_4
LB-Agar	LB, 1.5 % (w/v) agar-agar
Minimal Medium (M9)	0.2 x M9 salts, 19 mM NH_4Cl, 0.4 % (w/v) glucose, 1 mM $MgSO_4$, 0.0002 x trace metal mix, 0.01 x MEM vitamin solution (SigmaAldrich)
Trace metal mix	100 µM $FeCl_3$, 40 µM $CaCl_2$, 20 µM $MnSO_4$, 20 µM $ZnSO_4$, 4 µM $CoCl_2$, 4 µM $CuCl_2$, 4 µM $NiCl_2$, 4 µM Na_2MoO_4, 4 µM H_3BO_3
M9 salts	240 mM Na_2HPO_4, 110 mM KH_2PO_4, 1 % (w/v) NaCl

Table 12: Solution and buffers used in this work.

Solution/Buffer	Composition
6 x DNA loading buffer	10 mM Tris/HCl pH 7.4, 25 mM EDTA, 30 % (v/v) glycerol, 0.4 % (w/v) OrangeG
AP buffer	100 mM Tris/HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$
Blocking solution	TBST, 5 % (w/v) skimmed milk powder
Coomassie staining solution	30 % (v/v) ethanol, 10 % (v/v) acetic acid, 0.25 % (w/v) Coomassie R-250
Coomassie destaining solution	40 % (v/v) ethanol, 10 % (v/v) acetic acid
Congo red stock solution	0.2 % (w/v) Congo red, 1 % (w/v) NaCl, 80 % (v/v) ethanol, 0.001 % (w/v) NaOH
IB wash buffer A (*Ec*CsgA)	50 mM Tris/HCl pH 8, 1.5 M NaCl, 1 % (w/v) N-Lauroylsarcosine
4 x Lower buffer	1.5 M Tris/HCl pH 8.8
Lysis buffer I	50 mM Tris/HCl pH 8, 150 mM NaCl
Native buffer A (*Ec*CsgA)	50 mM KPi pH 7.2
Native buffer B (*Cs*gB)	5 mM KPi pH 3
Ponceau S staining solution	2 % (w/v) Ponceau S, 30 % (w/v) TCA, 30% (w/v) sulfosalicylic acid

Solution/Buffer	Composition
Y x SDS sample buffer	0.1 M Tris/HCl pH 6.8, 2 % (w/v) SDS, 0.2 % (w/v) Bromphenol blue, 20 % (v/v) glycerol, 5 mM DTT
SDS stacking gel (5 %)	0.8 mL Rotiphorese® Gel 30 (Carl Roth), 1.3 mL 4 x upper buffer, 3 mL MilliQ, 7.5 µL TEMED, 12.5 µL 25 % (w/v) APS
SDS separating gel (12 %)	6 mL Rotiphorese® Gel 30 (Carl Roth), 3.8 mL 4 x lower buffer, 0.15 mL 10 % (w/v) SDS, 5 mL MilliQ, 0.02 mL TEMED, 0.05 mL 25 % (w/v) APS
SDS separating gel (15 %)	7.5 mL Rotiphorese® Gel 30 (Carl Roth), 3.8 mL 4 x lower buffer, 0.15 mL 10 % (w/v) SDS, 3.55 mL MilliQ, 0.02 mL TEMED, 0.03 mL 25 % (w/v) APS
Solubilisation buffer A (*Ec*CsgA) , (*Cs*gB)	100 mM KPi pH 7.2, 150 mM NaCl, 8 M GdnHCl
1 x TAE buffer	40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, adjusted to pH 8.2 with acetic acid
1 x TBST	20 mM Tris/HCl pH 8, 150 mM NaCl, 0.05 % (v/v) Tween-20
Transfer buffer for Western Blotting	20 mM Tris pH 8, 192 mM glycine, 15 % (v/v) Methanol
4 x Upper buffer	0.5 M Tris/HCl pH 6.8, 0.4 % (w/v) SDS
Urea/SDS	8 M urea, 2 % (w/v) SDS

6.3 Microbiological and Molecular Biological Methods

Molecular biology methods, which were applied in this study, were adapted from standard procedures (Ausubel et al., 2007; Coligan, 2003; Sambrook and Russell, 2000) and are not described in detail unless essential modifications were made.

6.3.1 General Conditions of Bacterial Culture

Bacterial gene expression and plasmid amplification was done by freshly transforming with plasmid DNA by heat shock or electroporation. Followed by which transformed cells were incubated overnight on agar plates with an antibiotic, corresponding to the used plasmid. For liquid cultures an appropriate volume of culture medium was inoculated with a single colony. These cultures were incubated in baffled flasks under shaking (120 – 180 rpm, depending on number of baffles and ventilation). All bacterial cultures were grown at 20 - 37 °C. Plasmids were extracted from DH5α cells. BL21 (DE3) and T7 Express cells were subjected to bacterial culture for recombinant gene expression.

6.3.2 DNA Analytical Methods

DNA samples were analysed using agarose gel electrophoresis. DNA samples after PCR, modification by restriction enzymes and mutagenesis was visualized on 0.8 or 2 % (w/v) agarose gel in TAE buffer and supplemented with 5 % (v/v) ethidium bromide. Nucleic acid molecules are separated by applying an electric field corresponding to 5 V/cm for 60 min in TAE buffer to move the negatively charged molecules through the agarose matrix. The gel was documented under UV illumination at 254 nm.The yields and purity of isolated bacterial plasmid DNA, PCR products and modified DNA during cloning were checked spectrophotometrically by absorption at 260 and 280 nm. The DNA concentration was calculated from A_{260}, assuming that A_{260} = 1 is equivalent to 50 µg/mL DNA. With a ratio of A_{260}/A_{280} of 1.8 – 2 DNA was supposed to be pure.

6.3.3 Molecular Cloning

To establish the different constructs protein variants several cloning steps were used.
The enzymes, bacterial strains, plasmids and oligonucleotides, which were employed to this purpose, are summarised in **Table 3**, **Table 8, Table 9** and **Table 10.** If not stated otherwise, all reactions were carried out as described in the protocols of the manufacturers' instructions.
PCR was done using the Platinum *Pfx* Polymerase in order to amplify the gene of interest. Following the amplification the DNA fragment was purified by agarose gel electrophoresis followed by agarose gel extraction. The purified DNA as well as the recipient plasmids was digested with the corresponding restriction enzymes. Ligation was carried out after the dephosphorylation of linearized plasmids by Shrimp Alkaline Phosphatase to avoid self-ligation. Further in order to remove the modifying enzymes and remaining nucleotides, the PCR amplificates and the linearized plasmids were applied to the PCR purification kit. The ligation was performed by mixing plasmid- and target-DNA corresponding to molar ratio of 1:5 and heated to 90 °C for 5 min. This mixture was cooled down and incubated with T4 DNA Ligase at 16 °C overnight and finally 2 µL of the reaction volume were utilised for transformation. In order to modify plasmid-DNA selectively, QuikChange® Site-Directed Mutagenesis was performed using the *Pfu* Turbo Polymerase. All constructs were confirmed by sequencing in the department of Genome Analytics (HZI, Braunschweig)

using a 3100-Avant genetic analyser (ABI Prism) or by the company's GATC (Konstanz) and MWG (Ebersberg).

6.3.4 Generation of Plasmid Constructs to Obtain Fusion Proteins

A schematic representation of gene constructs encoding CsgB and CsgB truncation mutants are shown in **Figure C**.Gene constructs that encode CsgB-truncation mutants were cloned into the pET11d vector. Primers were designed to attach at the specific position on the full length gene from where the truncation was intended to start either on N-terminal or the C-terminal and amplified using the primer pair *csg*B-ΔN & ΔR5 fwd or rev and cloned into the given vector by *NcoI* and *Bam*HI.

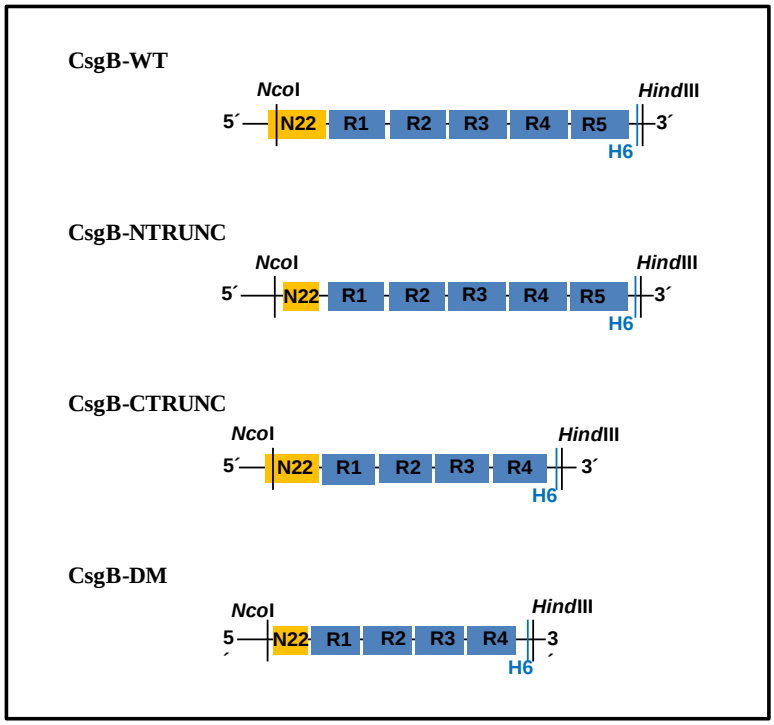


Figure C : Cloning sites for gene constructs encoding CsgA fusion proteins. CsgB-WT und CsgB truncation constructs with flexible N-terminal region shown in orange colour with CsgB repeating units (R1 – R5) are shown in blue.

7. Protein Production and Purification

7.1 Test Expression

Primary recombinant expression of novel gene constructs was carried out in small scale. From the freshly transformed plates single colony was inoculated into 20 ml liquid medium and cultures were grown to an OD_{600} of 0.6 – 0.8 at 37 °C. The cultures were induced with varying IPTG concentration, expression time and expression temperature. To determine the optimum parameters samples were drawn at several time points after induction samples were taken by V (µL) = $300/OD_{600}$. These samples were further analysed for soluble and insoluble expressions products using the SDS PAGE. Previous to SDS PAGE, insoluble fractions were solubilised with 8 M urea or 98 % (v/v) formic acid. Formic acid was removed using a speed vac at 60 °C for 1 h and the pellets were dissolved in SDS sample buffer as described in **7.6.4**.

7.2 Recombinant Protein Synthesis

For the recombinant expression of the protein, bacteria were cultivated as described in **6.3.1**. The main culture was always inoculated with approximately 2.5% volume of a pre-culture corresponding to the starting OD_{600} of 0.1, which was then grown at 37 °C. This main culture was grown to exponential phase (OD_{600} 0.6 – 0.8) at 37 °C and shifted to optimal expression temperature. Gene expression was induced by adding IPTG to a final concentration of 0.25 – 1 mM. Cells were harvested after 2 – 8 h induction followed by centrifugation (6000 g, 10 min, 4°C) and stored at -20 °C until further processing. The expression conditions are summarised in **Table 13.**

Table 13: Expression conditions.

Gene product	Host	[IPTG]	Induction temperature	Induction time
EcCsgA, CsgB, CsgB truncated constructs	T7 Express/DL39	0.5 mM	37 °C	2 h

7.3 Production of Isotope Labelled Protein

To achieve uniform isotope labelled protein, gene expression was carried out in M9 medium, which was enriched with an isotope labelled carbon and nitrogen source. Pre-cultures were grown in LB (CsgA) or in unlabelled M9 (CsgB) and later transferred into the M9 main culture. In case of CsgB, the pre-culture was harvested by centrifugation (6000 g, 10 min, and 4 °C) and re-suspended in unlabelled M9 medium. 2 L main culture were incubated until reaching an OD_{600} of 0.8, spun down as described and transferred again in 0.5 L fresh M9 enriched with $^{15}NH_4Cl$ as well as ^{13}C-glucose for double labelling followed by induction with 0.5 mM IPTG (Marley et al., 2001).

7.4 Cell Lysis

Cell pellets of CsgB-wt and CsgA-wt were thawed on ice and 1 g cells was suspended in 50 ml of 'Solubilisation buffer A' followed by incubation with mild stirring at at 4 °C for 48 hours or at room temperature corresponding to 25 °C for 14 hours.

7.5 Purification under Denaturing Conditions and Fibrillisation

Since most of the proteins, which were analysed during this study, were obtained as inclusion bodies after expression and cell lysis, they were purified under denaturing conditions comprising different conditions summarised in **Table 14** and illustrated below. The composition of the corresponding buffers is listed in **Table 12**.

Table 14: Purification under denaturing conditions.

Protein	IB purification	Solubilisation	IMAC	Desalting
*E*CsgA$_{21\text{-}151}$, *E*CsgB$_{21\text{-}151}$ *Cs*gB-ΔN *Cs*gB-ΔR5 *Cs*gB-ΔNΔR5	-	8 M GdnHCl → Sol. buffer A	Ni-Sepharose	Sephadex G25

Purification and Solubilisation of Inclusion Bodies

Inclusion bodies (IB) of CsgB-wt and CsgA-wt were solubilized in ‘Solubilisation buffer A’ followed by centrifugation (37000g, 30 min, 15 °C) to remove insoluble material. The supernatant containing the solubilized protein was then incubated with the Nickle-sepharose for further purification by affinity chromatography.

Affinity Chromatography

All proteins in this study carried a hexa histidine-tag and were applied to IMAC. The protein solutions were mixed either with equilibrated Ni-Sepharose or Ni-NTA matrix (Qiagen, 1 ml per 1 L bacterial culture) and incubated for 1 h at 4 °C to ensure binding to the matrix. Subsequently, the solid phase was packed into a column and immobilized proteins were washed threefold with 25 CVs per step (5, 10 and 20 mM imidazole in the corresponding solubilisation buffer). Finally proteins were eluted with 500 mM (5 CV) imidazole. Amount and purity of protein solutions were determined by SDS-PAGE (see **7.6.3** and **7.6.4**). Elution fractions of high yield and purity were combined.

Buffer Exchange and Fibrillisation

To allow fibril formation, the processed protein samples were transferred to native buffer lacking any denaturing agent and incubated accordingly.

<u>*Ec*CsgA, *Cs*gB, CsgB truncation mutants</u>:

After IMAC the proteinaceous sample were desalted using Sephadex G-25 medium in a Nap-5 or PD-10 column or a self-packed column (GE Healthcare) as follows:

The samples were spun down (20800 g, 10 min, RT), directly loaded on an equilibrated Nap-5 or PD-10 desalting column (GE Healthcare) and desalted as described in the manufactures protocol and summarised in **Table 15**. Proteins were eluted with either native buffer A (CsgA) or with native buffer B (CsgB) and fibrillised. CsgA was fibrillised in presence preformed $CsgB_{21\text{-}151}$ fibrils or with monomeric soluble CsgB-wt and CsgB mutants serving as seed or nucleator.

Table 15: Buffer exchange using Nap5 and PD10 columns (GE Healthcare).

Column	1st equilibration	sample volume	2nd equilibration	elution
Nap5	10 mL water, 10 mL buffer	500 µL	-	2 x 400 µL
PD-10	50 mL water, 50 mL buffer	0.5 – 1.5 mL	ad 2.5 mL buffer	4 x 850 µL
Sephadex G-25M gel filtration column	50 mL water, 100 mL buffer	1.5 – 2.5 mL	ad 2.5 mL buffer	10 x 1000 µL

Furthermore a self-packed column (Sephadex G-25 medium, 10/40, column volume ~ 30 mL) was used on an ÄKTA-FPLC system to desalt samples. 2 – 3 mL were spun down (20800 g, 10 min, RT) and loaded on the equilibrated column. The proteins were eluted with 1.5 column volumes of native buffer A with a flow rate of 2.5 mL/min.
Later experiments revealed that a filtration step of the solubilised and processed samples is necessary to obtain concentration dependent lag phases during fibril formation. Thus, proteinaceous samples were filtered before buffer exchange at room temperature by ultracentrifugation (50000 g, RT) using sorvall centrifuge..

$CsgB_{21\text{-}151}$

Samples of $CsgB_{21\text{-}151}$ were desalted as described above. To avoid the premature aggregation of $CsgB_{21\text{-}151}$, 5 mM KPi pH 3 was used for the desalting procedure. After the elution, the pH was adjusting the buffer concentration to 50mM KPi pH 7.2 and the samples were incubated for fibrillisation.

Preparation of Seeds

To ensure specific fibrillisation, protein samples of CsgA were supplemented with preformed fibrils of CsgB as seeds or monomeric soluble CsgB in case of nucleation studies. These fibrils were incubated up to four weeks (at least one week) at room temperature and prepared accordingly: Fibrils were sedimented (20800 g, 5 min, RT) and resuspended in native buffer A. The centrifugation and resuspension steps were repeated twice. Finally, the fibrils were added to freshly desalted protein samples.

7.6 Protein Analytical Methods

7.6.1 Physico-Chemical Parameters of the Studied Proteins

The physico-chemical parameters of the studied proteins are summed up in **Table 16** and were calculated using the programs PROTPARAM (www.expasy.org).

Table 16: Physico chemical parameters of the studied proteins.

	Position H6	**Number of amino acids**	**MW (kDa)**	**pI**	**ε_{280} (M^{-1} cm^{-1})**
*Ec*CsgA$_{21-157}$	C-term.	138	14.05	5.73	11460
*Ec*CsgB$_{21-157}$	C-term.	137	14.64	9.03	8940
*Cs*gB-ΔR5	C-term.	117	12.32	6.57	8940
*Cs*gB-ΔN	C-term.	119	12.63	9.90	5960
*Cs*gB-ΔNΔR5	C-term.	99	10.30	8.14	5960

7.6.2 Photometric Quantification of Protein Concentration

According to Lambert Beer law, the concentration of purified protein solutions was determined by detecting the absorption at 280 nm A_{280} against buffer. The molar extinction coefficient ε_{280} (**Table 16**) for each protein was calculated *in silico* (**7.6.1**).
Under denaturing conditions A_{280} was measured in spectroscopic pure guanidine hydrochloride. Under native conditions absorption was detected immediately after desalting avoiding fibrillisation.

7.6.3 NaDOC/TCA Precipitation

Proteins eluted in the guanidine hydrochloride containing solutions had to be precipitated by sodium deoxycholate and trichloracetic acid (Arnold and Ulbrich-Hofmann, 1999; Bensadoun and Weinstein, 1976) for analysis through SDS-polyacrylamide gel electrophoresis. Corresponding samples were diluted 1:10 in 0.1 % (w/v) NaDOC and mixed with TCA to a final concentration of 10 % (w/v). After incubation for 30 min at 4 °C the samples were spun down (20800 g, 30 min, 4 °C). The obtained pellets were washed twice by adding 500 µL chilled acetone, which was removed after centrifugation (20800 g, 10 min) and the pellets were air-dried to remove retained acetone. Dried pellets were dissolved in 20 µL SDS sample buffer and applied on a SDS gel.

7.6.4 SDS-Polyacrylamide Gel Electrophoresis

Discontinuous SDS-polyacrylamide gel (SDS PAGE) electrophoresis was used to analyse amount, purity and molecular weight of protein solutions. SDS PAGE electrophoresis was performed under denaturing and reducing conditions (Laemmli, 1970). The protein samples were mixed with an excess of SDS-sample buffer, boiled for 5 – 10 min at 95 °C and run on a suitable gel. The gels comprised of 5 % (w/v) acrylamide stacking gel and 12 or 15 % (w/v) acrylamide separating gel. The PAGE was performed at 40mA per gel for 28 and 33 min respectively. Furthermore commercial NuPAGE® Novex® 4-12% Bis-Tris Midi Gels (Invitrogen, 1.0 mm x 26 well) were run as described in the manufacturers' manual. If not stated, otherwise always 10 µL were loaded per lane. Finally, gels were stained with Coomassie R-250 and destained with corresponding solutions by boiling and shaking.

7.6.5 Transfer of Proteins to Membranes (Western Blot) and Immunodetection

Western blot was performed by transferring the protein from the SDS gels to the membrane. For immunodetection experiments protein samples in SDS-polyacrylamide gels were immobilised on polyvinylidene difluoride (PVDF) membranes (Immobilon P, Millipore). First, the membrane was activated in 100 % methanol (~ 10 s) and washed with transfer buffer. Freshly run gels were equilibrated in transfer buffer as well (15 min). Following, the gel was placed onto the membrane, and both placed between two layers of soaked Whatman paper onto the anode of a semi-dry blot apparatus. The transfer was carried out for 40 min at 15 V.After western blotting unspecific binding sites on the PVDF membrane were saturated for 1 h in blocking solution and washed in TBST (5 min). Antibodies were diluted in TBST/0.5 % (v/v) blocking solution. The membrane was incubated with the primary antibody (**Table 4** in **6.1.1**) overnight and after washing with TBST (3 x 5 min) additionally incubated for 2 h with the secondary antibody (**Table 5** in **6.1.1**). Finally, the membrane was washed (TBST, 3 x 5 min) and immobilised proteins were detected by BCIP/NBT (Promega) or Enhanced Chemiluminescence (ECL, LumiLight, Roche).

7.6.6 N-Terminal Sequencing

For N-terminal sequencing experiments protein samples in SDS-polyacrylamide gels were immobilised on polyvinylidene difluoride (PVDF) membranes (Immobilon P, Millipore). Proteins, which were transferred to a membrane as described in 7.6.5, were excised after visualising with Ponceau S staining solution and sequenced by automated Edman degradation (Edman and Begg, 1967). N-terminal sequencing was carried out by Rita Getzlaff or Beate Jaschok-Kentner (HZI, Braunschweig) using a 494A HT Protein Sequencer (Applied Biosystems).

7.6.7 Mass Spectrometry

Mass shifts were analysed to confirm the size of full length or the truncation mutant recombinant protein expression as well as the modification of amino acid. MALDI-TOF-MS was used to verify the mass and intactness of recombinant produced proteins. The fibrils (approximately 25 - 50 µg) were dissolved in 98 % (v/v) FA, loaded onto a prespotted anchor chip (PAC) target with a α-cyano-4-hydroxycinnamic acid matrix and dried at RT. To avoid N-formylation the FA containing solutions were processed as fast as possible. The molecular masses were determined in the positive-ion mode on a Bruker Ultraflex time-of-flight mass spectrometer (Bruker Daltonics GmbH). All experiments were performed by Manfred Nimtz, Anja Meier and Undine Felgenträger (HZI, Braunschweig).

8. Protein Structure Analysis

8.1 Secondary Structure Analysis

The secondary structure of the proteins was determined using biophysiocal methods like CD and FT-IR spectroscopy. In CD experiments the differential absorption of left and right circular polarised light by chiral molecules is recorded. FT-IR is based on measuring molecular bond vibrational frequencies. Thus, both techniques provide information about the content of β-sheets, α-helices and loops of a protein structure.

Circular Dichroism

For CD experiments, protein samples were either desalted in to the CD suitable buffer (50 mM KPi pH 7.2 or water) followed by fibrillization or specific aggregates was washed with the respective buffer. If not stated otherwise, protein concentration was adjusted to 1 mM peptide bonds and samples were applied to cuvettes with a path length of 1 mm (Hellma). Spectra were taken using a spectropolarimeter JASCO J-810 and accumulated (2 – 3 spectra). Measurements were performed in far UV (190 – 260 nm) range in continuous scanning mode with a resolution of 0.2 – 0.5 nm at 4 °C (soluble proteins) or 20 °C (fibrils).

Fourier Transform Infrared Spectroscopy

FT-IR spectroscopy was done using a Bruker Tensor 27 system (Bruker Optics) equipped with a BioATR cell (ZnSe single crystal), a mid-IR source (4000 – 400 cm^{-1}) and a KBr beam splitter. The protein samples were applied as fine suspension with a protein concentration of 1 – 2 mg/mL. In preparation of the FT-IR experiments, fibrils were washed several times with water by repeating centrifugation (20800 g, 5 min, RT) and re-suspension steps. Measurements were carried out in absorbance mode with the following settings: sample and background scan time was 200 scans, optical filter was open and aperture was set to 6 mm, detector setting was given by LN-MCT Photovoltaic and preamp gain A.

8.2 Probing the Amyloid fold as Tertiary/Quaternary Structure

The specific aggregation into the amyloid fold was visualized by different imaging, staining and biochemical methods. Usually specific aggregates of amyloid fibrils are

obtained as long fibrous structures; electron microscopy allows differing between unspecific and specific aggregates. Most amyloids can be stained by the amyloid specific dyes like CR and ThT. The binding of CR is detected by the characteristic green birefringence under cross-polarised light. The interaction of ThT with amyloid fibrils results in an increased and robust ThT fluorescence at 485 nm.

8.3 Electron Microscopy

The greatest challenge to examine biological material in a TEM is the non physiological condition to which specimen must be exposed. First, since specimens are examined in a high vacuum, they have to be dry. Secondly, the image contrast in a TEM results from electron scattering and contrast correlates directly with the amount of scattered electrons. The higher the atomic number, the more electrons are scattered and the greater the contrast. Unfortunately, most biological elements like carbon, hydrogen, oxygen or nitrogen don't scatter electrons to gain a good contrast. In order to improve contrast, heavy metals have to be added during the preparation steps. Thirdly, electrons have a limiting power to penetrate the specimen. This requires a preparation of a very thin specimen or slicing the specimen into thin sections of 50-100 nm. In summary, the preparation of biological specimen while retaining the native structural morphology was, is and will be a challenge for TEM imaging. Here the main method used and described in detail is negative-staining.

Negative-Staining Methods

Negative staining is the most rapid, reproducible and easy to perform method for analyzing bacteria, phages or enzyme molecules in the TEM. Since negative staining involves deposition of heavy atom stains like uranyl acetate on the specimen with subsequent air-drying, structural artefacts such as flattening are commonly observed. It is assumed that the heavy metal salt solutions occupy hydrated regions of the specimen and penetrate into cavities within the sample. The TEM image represents a projection of the sample, which is electron translucent whereas the surrounding areas of the sample are grey to dark. Nowadays, cryo-electron microscopy is currently regarded as the best method to view small native specimens after flash freezing the samples in a thin film of vitreous ice and imaging without staining. Nevertheless, negative staining is the method of choice to characterize bacteria and to analyse surface appendages like flagella, pili or fimbriae.

Negative staining with carbon film

Specimen and supporting films require a mechanical support in the form of a metal grid for imaging in the TEM. The grid itself exhibits an electron opaque component, the bars, and an electron translucent part, the open area. The number of grid bars and percentage of open area can be varied depending of the grid types used; 100-400 mesh (lines/in) grids are commonly used. Most EM grids are made of copper because they are the cheapest and their non-ferromagnetic character results in minimal distortion of the magnetic field of the objective lens. In addition, the copper mesh also conducts heat away from the support film and helps prevent thermal expansion and hence movement of the specimen under the electron beam can be minimized. For immune labelling studies nickel grids are used since copper grids develop copper rust with time when incubated with high salt buffers, like PBS. The following three negative staining solutions containing heavy metal solutions are commonly used:

• Aqueous 0.5%-4% uranyl acetate solution, pH around 4.5, if pH needs to be adjusted to the more basic side add 0.1 M KOH carefully, above pH 5.0 non soluble $UO_2(OH)_2$ is formed, by adding 1 mM EDTA the pH can be raised to 7.5 but the grain size of the stain will increase.

• Aqueous 0.5%-3% phosphotungstic acid solution or corresponding Na and K salts, pH is adjustable with 1 M KOH or NaOH.

• Aqueous 1%-10% ammonium molybdate, for pH adjustment titrate with 0.1 M KOH, NaOH or ammonia.

• Pipette 50 µl droplets of the sample, washing and staining solutions on a clean piece of Parafilm™. Take care that a pronounced convex meniscus is always visible. For washing solutions, TE-buffer (20 mM TRIS, 1 mM EDTA, pH 6.9) and distilled water can be used.

• Cut a square of 2 mm x 2 mm from the carbon coated mica and hold the piece with sharp, pointed forceps at one end of the piece of mica.

• Introduce the carbon coat piece of mica at a 45° angle slowly into the sample solution. The carbon film will float off the mica and particles will start to adsorb to the carbon film. Part of the carbon film still has contact to the mica and is held in place by the forceps. The floating time of the carbon film on the sample solution depends on the number of particles in the solution.

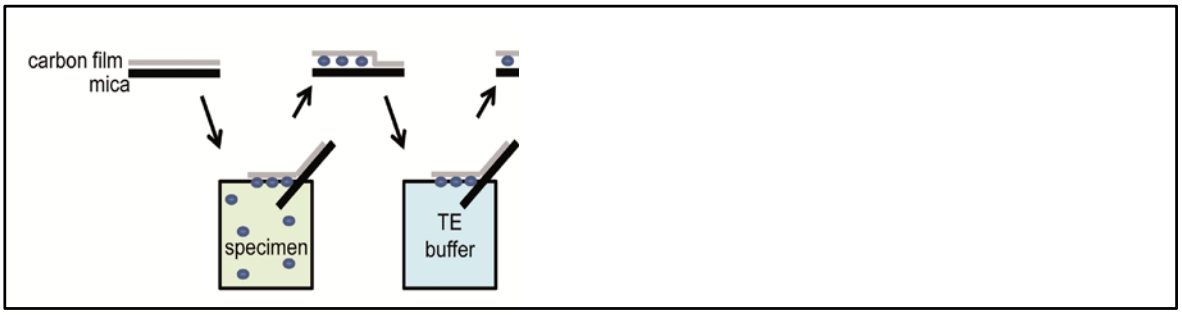


Figure B: **Preparation of grid** (Manfred Rohde)

Generally, the length of the adsorption time (normally 10-60 seconds), depends on the number and size of the particles in the sample solution. For bacteria the sample solution should be slightly turbid; for phages it is recommended to use 10^7-10^9 phages/ml and for proteins a concentration of 50-100 µg protein/ml is sufficient. To obtain reasonable results for the number of particles per viewing field in the TEM the trial-and-error method applies.

• Remove the piece of mica slowly from the sample solution and the carbon film will fall back into its original position. Remove excess sample solution by blotting on filter paper, but do not blot dry.
• Using the same procedure as described above transfer the carbon film onto TE-buffer and distilled water, always remove excess fluid by blotting on filter paper.
• Completely float the mica off onto the staining solution by moving the piece of mica under the surface of the staining solution drop, open the forceps and the piece of mica will fall to the bottom, and the carbon film floats on the staining solution.
• Pick up the carbon film with the adsorbed specimen using a 300 mesh copper grid. The grid should be cleaned before with acetone. Since grids have the tendency to be hydrophobic it is recommended to put the grid with the rough side on a drop of distilled water, leave for a few seconds and blot on a filter paper and immediately put the grid onto the carbon film. Slightly press on the grid with the forceps to get better contact of the film to the grid and to remove the occasionally build air bubbles between carbon film and the grid.
• Pick up the grid perpendicular from the staining solution drop, turn it around and blot the staining solution from the grid with filter paper. Put the filter paper at the edge of the grid and let the solution creep up the filter paper. If the staining solution does not flow anymore into the filter paper remove the filter paper immediately. Using this procedure one usually gets a gradient of shallow stain and deeper stain on the carbon film. Blotting of staining

solution is the most crucial step in the entire negative staining procedure and needs a bit of training to reach the desired staining thickness.

• Air-dry the grid with the attached carbon film with the help of a warm light globe by placing the grid under the globe at a distance of 1-2 cm for 10-15 sec.

• Examine by TEM.

Summary

In order to analyse fibrils by electron microscopy, appropriate samples were deposited on a carbon-coated grid and stained negatively by a method adapted from (Valentine et al., 1968).

Appropriate samples were negatively stained with 4% (w/v) aqueous uranyl acetate using 300 mesh nickel grids (Plano). To coat these grids, thin carbon support films, approximately 10 nm thick, were prepared. Carbon was sublimated from a carbon thread on to freshly cleaved mica (SCD500, Bal-Tec). Finally, the negatively stained samples were examined in a Zeiss 910 transmission electron microscope (Zeiss) with an acceleration voltage of 80 kV and at calibrated magnifications. Images were recorded digitally with a Slow-Scan CCD-Camera (ProScan, 1024x1024) with ITEM-Software (Olympus Soft Imaging Solutions). All experiments were performed by Manfred Rohde and Heinrich Lühnsdorf (HZI, Braunschweig)

8.4 Thioflavin-T Fluorescence Studies

The aggregation kinetics of CsgA and CsgB were determined by increase in Thioflavin dye binding monitored using Tecan-Platereader. In time-resolved ThT studies 200 µL of fresh desalted amyloidgenic proteins (2 – 16 µM) were mixed in to a 96 well plate (Corning, Flat Bottom, Non-Binding Surface, Non-Sterile, Black Polystyrene) and subsequently incubated in presence of saturating concentration of ThT. The ThT fluorescence intensity was monitered at 485 nm after excitation at 442 nm every 10 min using a Tecan-Platereader. The measurements were performed at 25 °C. For both, excitation and emission, a band width of 5 nm was used and multiple readings were obtained per well (5 x 5, border 1000 µM). The homogeneity of the samples was ensured by shaking the plate before every measurement for 2 s (amplitude 2 mm). The gain (100 – 150) and the z-position (~ 21000 µm) were adjusted manually before starting the assay using preformed fibrils.

8.5 NMR spectroscopy

8.5.1 General Measurement Conditions

All NMR experiments were carried out at 25 °C using a Bruker Avance III 600 spectrometer equipped with four radio-frequency channels and a 5 mm Z-axis gradient triple-resonance cryo-probehead. The offset that represses the water signal and the pulse length were optimised for every individual dataset with the appropriate samples.

Amyloid fibrils were collected by centrifugation (20800 g, 4 min, RT) and solubilised in ~ 500 µL aprotic DMSO (perdeuterated, d_6-DMSO) containing up to 0.1 % (v/v) TFA (deuterated, d_1-TFA), and the samples were placed in a 5 mm NMR tube (Norrel). For 1D- and 2D-experiments ~ 20-25 mg wet pellet of fibrils was processed, for 3D-experiments up to 20-30 mg were used. Under native conditions 50 mM KPi pH 7.4 was used as buffer and complemented with 10 % (v/v) D_2O prior measurements. The protein concentration was ~ 50 µM.

All experiments and most important parameters are summed up in **Table 17**; sample preparation is shown in **Table 18**. The spectra were processed and analysed using the programs PROSA (Guntert et al., 1992) and CARA (Keller, 2004).

Table 17: NMR experiments carried out in this work and their parameters.

SW: sweep width (spectral width), TD: time domain (number of obtained data points), NS: number of scans

Experiment	Obtained correlation	SW (Hz)	TD	NS	pulse programme
		1D-experiments			
1D-^{1}H	^{1}H	8417.5	2048	128	zgpr
		2D-experiments			
ctHSQC	^{1}H(i), ^{15}N(i)	4807.7 1338.3	2048 192	2	NHSQC-ct (Mori et al., 1995)
fHMQC (Luhrs et al., 2005; Ritter et al., 2005)	^{1}H(i), ^{15}N(i)	4807.7 1338.3	1024 128	2	fhmqc, time optimised HMQC (modified from (Bax et al., 1983)
		3D-experiments			
HNCA	$^1H^N$(i), ^{15}N(i), $^{13}C_\alpha$(i), $^{13}C_\alpha$(i-1)	8417.5 1338.3 4830.6	4096 62 196	16	hncapwg3d (Grzesiek and Bax, 1992)
HN(CO)CA	$^1H^N$(i), ^{15}N(i), $^{13}C_\alpha$(i-1)	8417.5 1338.3 4830.4	4096 62 196	16	hncocapwg3d (Grzesiek and Bax, 1992)
HNCACB	$^1H^N$(i), ^{15}N(i), $^{13}C_\alpha$(i), $^{13}C_\alpha$(i-1), $^{13}C_\beta$(i), $^{13}C_\beta$(i-1)	8417.5 1338.3 11312.2	2048 64 192	32	hncacbgp3d (Wittekind and Mueller, 1993)

Experiment	Obtained correlation	SW (Hz)	TD	NS	pulse programme
HNH-NOESY	$^{1}H(i) \xrightarrow{NOE} {}^{1}H(j), {}^{15}N(j)$	8417.5 8403.4 1338.3	2048 384 72	8	HNH_noesy (Marion et al., 1989)

Table 18: Sample preparation for NMR experiments.

Protein	Experiment	wet pellet	[TFA]	volume
*Ec*CsgA	1D & 2D	20 – 25 mg	0.1 % (v/v)	520 µL
	3D	35 mg		
*Cs*gB	1D & 2D	~ 20 mg	0.1 % (v/v)	500 µL
	3D	20-30 mg		

8.5.2 Sequence Specific Resonance Backbone Assignment

For sequence specific backbone assignment of curli protein CsgB, ^{13}C,^{15}N-labelled fibrils were subjected to triple-resonance-experiments HNCACB (Wittekind and Mueller, 1993), HNCA and HN(CO)CA (Bax and Hofstee, 1991) as well as HNH nuclear Overhauser enhancement spectroscopy (Diercks et al., 1999) experiments to achieve the sequence-specific resonance assignment of the backbone H^N/N cross-peaks. HNCACB, HNCA and HN(CO)CA enabled the correlation of specific backbone amide proton and nitrogen resonances with the corresponding alpha- and beta-carbon resonances of that particular amino acid. These experiments resolve the proton/nitrogen correlations in the same fashion as [^{15}N,^{1}H] correlation experiments with the third dimension of the spectra showing ^{13}C-chemical shifts of carbon resonances summarised in **Table 17** and **Figure A**

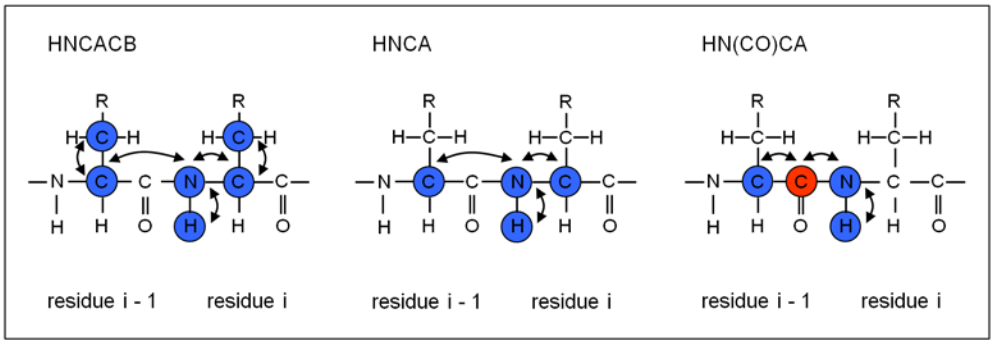


Figure A: Magnetisation transfer during the triple-resonance experiments HNCACB, HNCA and HN(CO)CA. (Wittekind and Mueller, 1993) (Bax and Ikura, 1991) The nuclei, whose frequencies are detected, are indicated in blue. The nuclei, which only transfer the magnetisation, are shown in red. The arrows indicate the magnetisation direction.

Due to different magnetisation transfer pathways, HNCACB spectra display ^{13}C-chemical shifts of the C_α and C_β resonances of the same residue as the backbone amide and the preceding residue in one strip. HNCA spectra display C_α shifts of a given residue and the predecessor, while HN(CO)CA spectra exploit the C_α shift of the previous residue only.

C_α- and C_β-resonances are differentiated by their opposite sign. Resonances of the same residue as the backbone amide and the preceding residue are distinguished by their intensities. Based on this information, peaks are aligned establishing that the corresponding amide pairs are adjacent in the sequence. The C_α- and C_β-chemical shifts were compared to the average of chemical shifts for each amino acid deposited in the BMRB (www.bmrb.wisc.edu) since the protein was in random coil state. This allowed the assignment of the strips to the protein sequence. The assignments were further confirmed by HNH nuclear Overhauser enhancement spectroscopy experiments. Each strip (3rd dimension) of a NEOSY-HSQC spectrum contains NOEs from one backbone amide group to all protons close by in this case the adjacent amino acid. Solubilising the fibrils led to unstructured proteins hence the obtained signals corresponded only to short and medium range NOEs enabling the alignment of adjacent backbone amides.

8.5.3 H/D Exchange NMR Experiments

^{15}N labelled fibrils of *Ec*CsgA, CsgB-wt and CsgB mutants were used for H/D exchange studies of the backbone amides (Hoshino et al., 2002; Li and Woodward, 1999). For HDX experiments the protein samples were fibrillized at a very low concentration corresponding to approximately 30 µM and washed with 0.1 % SDS to wash away the non specific

aggregates followed by washing with 'native buffer A' to remove SDS. Since the ordered fibers are resistant to the SDS wash only the non specific aggregates were dissolved by SDS. The above procedure enhanced the population of specific ordered fibers.

The exchange reaction was started by sedimenting the fibrils at 20800 g for 4 min, washed in corresponding native buffer comprising D_2O as the solvent, sedimented again and re-suspended in the same buffer for incubation up to 12 weeks at room temperature. Each aliquot contained around 20 – 25 mg of fibrils as wet pellet. Hydrogen exchange was quenched at suitable intervals (15 min, 1 h, 1 d, ..., 20 w) by sedimenting the fibrils (20800 g, 4 min, RT) and freezing the pellet in liquid nitrogen.

A series of up to 80 two-dimensional [^{15}N,^{1}H] correlation spectra each spectra corresponding to 4.5 mins were recorded directly after solubilising the fibrils in ~500 µL deuterated dimethyldisulphoxide (d_6-DMSO) containing deuterated trifluoric acid (d_1-TFA). The amount of residual D_2O was about 4 % (wet pellet of fibrils). Residues that display a fast exchange in the fibrils and residues with high intrinsic exchange rates in DMSO result in absent peaks in the [^{15}N,^{1}H] correlation spectrum. To differentiate them, a second series of two-dimensional spectra were measured after addition of 4 % (v/v) H_2O.

Next, the peak volumes within a series of NMR spectra were integrated for each amino acid and extrapolated to t_{DMSO} = 0 min. The extrapolated peak volume displayed the resonance intensities corresponding to the incubation time of fibrils in D_2O (**Equation 2**). To determine the specific exchange rates these intensities were fitted monoexponential to the exchange time (**Equation 2**).

$$I_t = \left(I_0 + I_\infty\right) \cdot e^{-k_{ex} \cdot t} + I_\infty$$

Equation 2: Exponential decay. I_t gives peak volume at certain time, I_0 peak volume at t = 0, I_∞ at t_∞, k_{ex} exchange rate and t certain time.

The data were analysed by using the programs PROSA (Guntert et al., 1992) and CARA (Keller, 2004), and a special written program in Visual basic in combination with Microsoft Excel (Luhrs et al., 2005)

9. Results

9.1 Chapter 1 - Characterization of curli nucleator protein CsgB

9.1.1 Abstract:

Nucleation is the rate-limiting step of amyloid propagation and its nature remains poorly understood. Nucleation is an important step in the formation of functional amyloids. The nucelator makes sure that the right kind of template is provided for fiber growth thus avoiding random seed formation which would lead to heterogeneous fibers and possibly formation of toxic intermediates. Contrary to the *in vitro* condition the minor curli subunit protein CsgB, is required for curli polymerization *in vivo* (Agrawal et al., 2011; Aguzzi, 2008) (**Curli specific proteins**). The aim was therefore to elucidate the role of CsgB in template-mediated nucleation of CsgA. *In vitro* formed CsgB fibrils were characterized by CD, FTIR, which indicated the presence of beta sheet rich secondary structure. We characterized the sequence specific secondary structure of CsgB through solution state NMR using quenched H/D exchange method. The study revealed higher level of heterogeneity among the fibers indicating the presence of different conformations. The secondary structure revealed the presence of four pairs of beta strand within the fiber core comprising of residues 45-132 which fulfill specific roles (**General Amyloid features)** in directing amyloid formation. Interestingly, the N-terminal flexible region 21-44 showed an additional single beta strand and the C-terminal 133-151 showed complete protection with higher level of heterogeneity. These results show a difference in the N and C terminal secondary structure of CsgB contrary to the predicted model (White et al., 2001). Based on these results nucleation model of a CsgB showing skewed (neither parallel nor intersecting) and overlapping organization of beta sheet elements providing a sequence selective interface for the nucleation of CsgA was hypothesized (**Figure 24**). Analysis of CsgB through electron microscopy and kinetic studies with the amyloid binding dye thioflavin T revealed faster aggregation kinetics compared to the major subunit protein CsgA. These results confirmed amyloidogenic property of CsgB as nucleator. We developed a new nucleation assay which led to the novel finding that monomeric CsgB can act as a nucleator contrary to the recently published belief that only fibrillar or oligomeric forms of CsgB can act as a nucleator *in vivo* (Ahmad et al., 2009).

Despite 30% identity in the amino acid sequence between CsgA and CsgB and in silico molecular modeling predicting a similar cross beta-strand structure (Collinson et al., 1999b; White et al., 2001), even though CsgA and CsgB show different functions in vivo. This part of the results includes two main aims 1) to devise a new purification protocol to purify full length CsgB for structural study using NMR and biophysically characterizing the protein. 2) to design a novel and robust protocol to monitor the folding or aggregation kinetics of CsgB and establish the nucleation assay of CsgA by CsgB to mimic the in vivo physiological condition.

9.1.2 Starting Point of CsgB Structure Elucidation

CsgB was initially purified by the Chapman group with the secretion peptide attached to the protein (**Figure 9)** hence was affinity purified from cell free supernatants. However, attempts to purify WT CsgB were unsuccessful due to low yields. To overcome the problem of low yield, a truncation mutant missing 19 amino acids from its C-terminus was used for further study (Hammer et al., 2007b). To elucidate the molecular mechanism of curli fiber nucleation process it was necessary to use the full length protein in order to emulate in vivo conditions.

9.1.3 CsgB-wt (full length) protein purification

Full length CsgB-wt lacking the secreation peptide was cloned in a pet11d vector (construct by A.Eberth) and was affinity purified under denaturing conditions to obtain pure and monomeric full length CsgB with approximate yields of 10mg per liter of LB as described in material methods. One of the major problems in using full length CsgB was the uncertainty of obtaining monomeric form of the protein due to its faster aggregation kinetics. The folding was initiated by desalting of CsgB into the native potassium phosphate buffer at pH 7.4 to remove the denaturant GuHCl. This resulted in rapid precipitation of the major portion of the protein on the G-25 sephadex gel filtration column resulting in the loss of protein. The process of rapid precipitation also caused the remaining monomeric CsgB to aggregate faster making it difficult to purify CsgB as a monomeric protein. In order to study folding kinetics and nucleation, the monomeric form of CsgB is an absolute prerequisite. To solve the above mentioned problems a closer look at the amino

acid sequence of CsgB revealed higher percentage of charged amino acids mainly in the C-terminal region comprising of residue 133-151 and a very high calculated pI of 9 which hinted that the folding of CsgB may depend on the ionic interaction which led to its faster aggregation kinetics. The pI value can affect the solubility of a molecule at a given pH, hence to gain a control over the folding process protein was desalted in to native buffer at low pH to get a monomeric soluble CsgB. Aggregation of CsgB was started by swiftly bringing the protein to higher pH, hence varying the pH acted as a switch to control the aggregation kinetics.

9.1.4 Production of monomeric soluble CsgB-wt using pH switch

Higher percentage of charged amino acids in CsgB led us to investigate its aggregation property at low pH. CsgB was freshly renatured by desalting from guanidine buffer to 5 mM potassium buffer at pH3 and monitored its binding to thioflavin T. In contrast to CsgA (Dueholm et al., 2011), aggregation of CsgB at pH3 was drastically reduced (**Figure 16b)** indicating that self folding and/or aggregation of CsgB depend mainly on ionic interaction between the residues. The aggregation of CsgB was faster at higher pH (**Figure 16a**) showing the amyloidgenic property can be controlled using pH as switch.

9.1.5 CsgB-wt forms amyloid-like fibers in vitro

In order to determine the amyloidogenic property of CsgB, affinity purified protein in GuHCl stored at 4° C for not more than a week was used. Immediately after desalting to pH 3 buffer, CsgB-wt was SDS soluble and migrated to its predicted molecular weight in an SDS-PAGE gel (Fig. 14 a) and was confirmed by western blot (Fig. 14 b) using polyclonal CsgB antibody. The denaturant was removed by desalting using sephadex G25 gel filtration column in to 5 mM potassium phosphate buffer at pH 3, immediately followed by changing the buffer conditions to 50mM potassium phosphate pH7.4, using the 1M stock of potassium phosphate pH 7.4. Increasing the pH and incubation at room temperature showed the formation of precipitate instantly this was then allowed to aggregate for 24 hours. Interestingly CsgB-wt showed visible precipitate instantly after increasing pH even at low concentration of 10 µM compared to CsgA, indicating the faster aggregation property of CsgB. These precipitates were analyzed by various biophysical methods to characterize and to test the amyloidogenic properties of full length CsgB-wt.

9.1.6 Morphology of precipitates

The ultrastructure of CsgB-wt aggregates was investigated by transmission electron microscopy (TEM). Freshly purified CsgB was desalted in pH 3 buffer which was then adjusted to pH 7.4 using 1M potassium phosphate stock solution. TEM analysis of purified 30 µM CsgB-wt revealed the presence of highly ordered, amyloid-like fibers (Fig. 14 c).

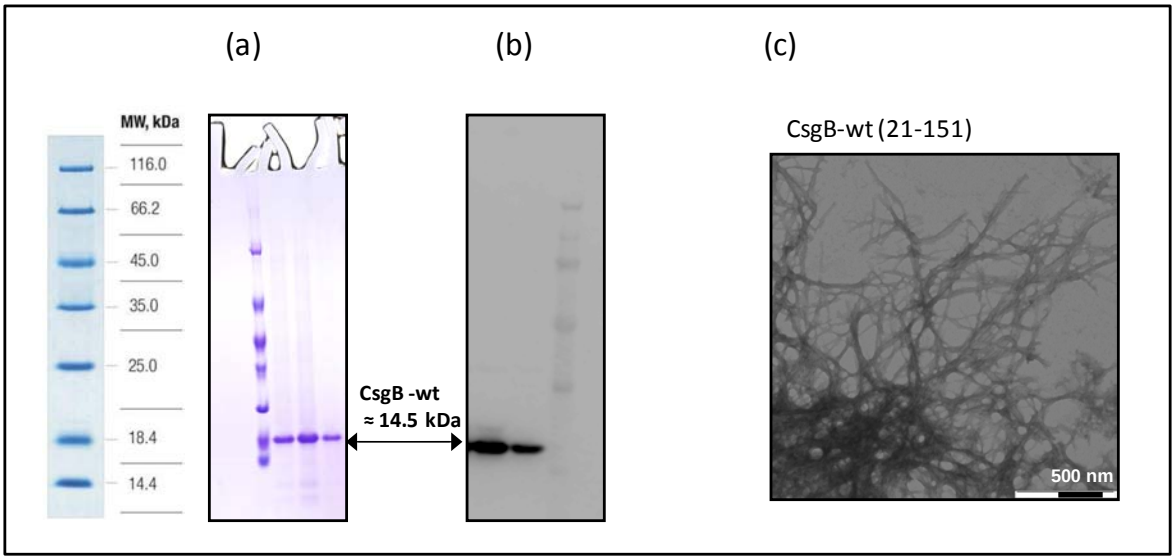


Figure 14: Biochemical property of CsgB.

(a) Coomassie stain SDS-PAGE of CsgB-wt samples. The molecular weight of the protein is 14.5 kDa, but runs at 18 kDa on the gel. Molecular size markers (in kilo daltons) are indicated on the left. **(b)** Western blot of CsgB-wt **(c)** Negative-stain EM micrograph of purified CsgB-wt. Magnification indicated on the right corner (500nm).

9.1.7 Secondary structure analysis CsgB-wt

The structural changes that occur during CsgB-wt polymerization were measured using Circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR). Freshly purified CsgB was desalted in buffer with pH 3 which was then adjusted to pH 7.4 using 1M potassium phosphate stock solution. The CD spectra of freshly renatured 10 µM CsgB-wt displayed a trough at 198 nm, which is characteristic of proteins that adopt a random coil structure (**Figure 15c** – red curve). The transition from the random coil to the beta sheet started within ten minutes of increasing pH. Following incubation of CsgB-wt at room temperature for 24 hours the CD spectrum revealed a trough at 218 nm, indicative of a β-sheet rich conformation (**Figure 15c** – green curve).

The secondary structure of the protein was further determined using BioATR–FTIR. The FTIR spectra showed amide I band between 1700 cm−1 and 1600 cm−1. Spectra of freshly purified CsgB-wt displayed Amide I maximum 1647 cm−1 (**Figure 15a** red curve). Following incubation of CsgB-wt at room temperature at pH 7.4 for 24 hours the maximum of the amide I shifted to 1620 cm−1 with an additional low-intensity peak at 1659 cm−1 (**Figure 15a** green curve). This transition was fast and started within the 10 minutes of increasing the pH (**Figure 15a** red curve) displaying the characteristic of a cross beta sheet structure (Zandomeneghi et al., 2004). The FTIR spectrum was similar to CsgA (**Figure 15b**) indicating CsgB of having similar secondary structure.

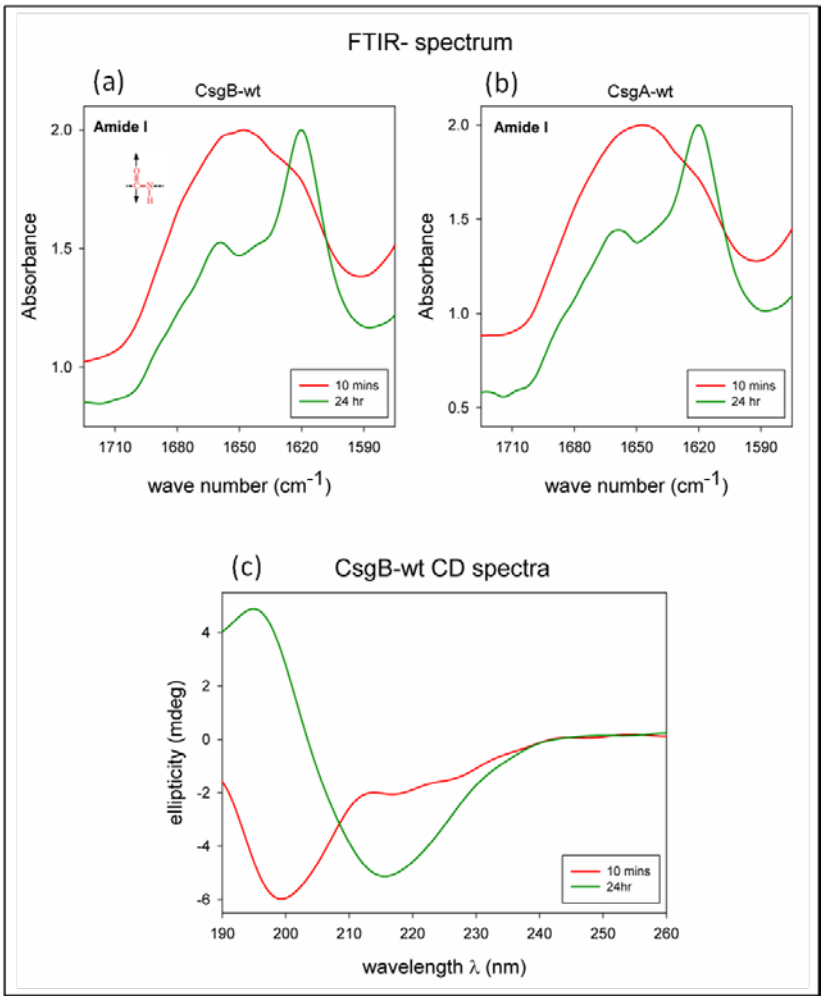


Figure 15: Biophysical characterization of CsgB.

(a) FT-IR spectra showing the amide I band of aggregates (1 mg/ml) formed by CsgB-wt samples. Aggregation kinetics analyzed by FTIR at zero hour (red curve) and 24 hour (green curve). **(b)** FT-IR spectra showing the amide I band of aggregates (1 mg/ml) formed by CsgA-wt samples. Aggregation kinetics analyzed by FTIR at zero hour (red curve) and 24 hour (green curve). **(c)** Far-UV-CD spectra of aggregates in ddH_2O formed by CsgB-wt. Aggregation kinetics analyzed by CD at zero hour (red curve) and 24 hour (green curve).

9.1.8 CsgB-wt aggregation kinetics

The amyloidogenic property of CsgB-wt was analyzed by binding with amyloid-specific dye, thioflavin T. The aggregation kinetics was monitored through an increase in fluorescence upon dye binding. Freshly purified varied concentration of CsgB (1, 2, 3,5,7,9 µM) was mixed with thioflavin T simultaneously increasing the pH to 7.4. The thioflavin T binding showed increased fluorescence over time in a concentration dependent manner with very less or nearly no lag phase (**Figure 16a**), in contrast to the aggregation kinetics of CsgA where it showed about 200-300 mins lag phase under the same buffer conditions (**Figure 17a** black curve). CsgB showed faster aggregation kinetics thereby indicating its role as a potent nucleator in the curli biogenesis. In contrast, at pH 3 CsgB showed no or little aggregation over time (**Figure 16b**).

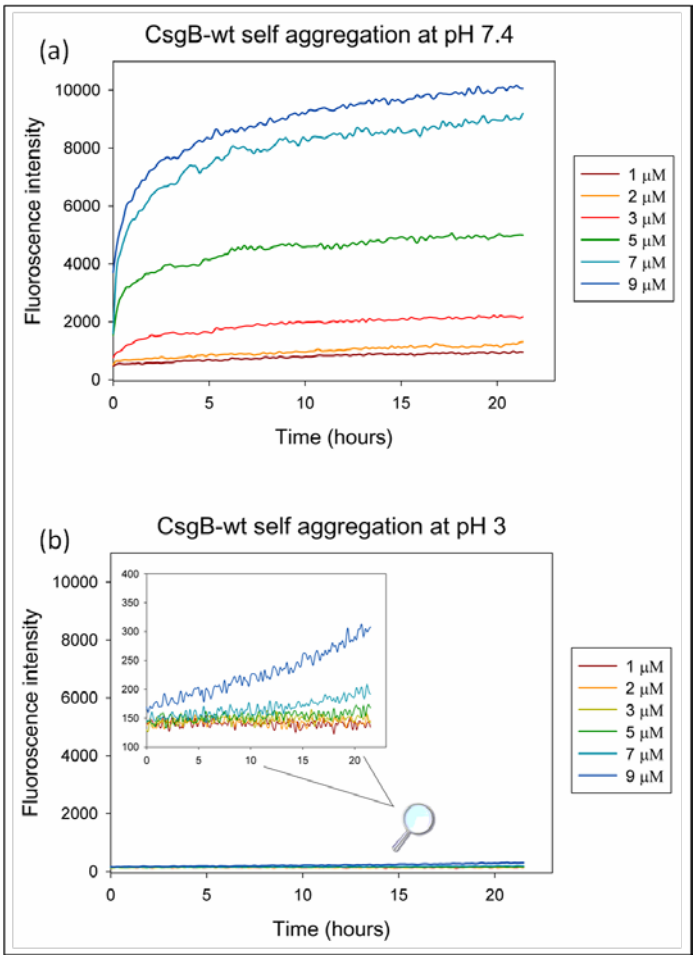


Figure 16: Aggregation kinetics monitored by Thioflavin T binding.

(a) Fibrillation of CsgB-wt followed by ThT at pH 7.4. Increase in ThT fluorescence depending on time and protein concentration (1, 2, 3,5,7,9 µM) **(b)** Fibrillation of CsgB-wt followed by ThT at pH 3. Little or no increase in ThT fluorescence depending on time and protein concentration (1, 2, 3,5,7,9 µM).Inside is the zoomed picture.

The CsgB aggregates are β-sheet-rich fibers that interact with the amyloid specific dyes thioflavin T. These biochemical and biophysical features are consistent with the hypothesis that CsgB-wt forms amyloid-like aggregates.

9.1.9 CsgB-wt fibers can seed CsgA in vitro

The biochemical and biophysical analysis of CsgB-wt demonstrated its amyloidogenic properties. The next question was if CsgB-wt fibers provided a seeding surface for CsgA polymerization. CsgA transition into an amyloid fiber in vitro is characterized by distinguishable lag, growth, and stationary phases. The lag phase can be shortened by the addition of preformed CsgA fibers (Agrawal et al., 2011). When mixed with thioflavin T, the relative fluorescence units of an 8μM solution of freshly purified CsgA increased after a 300 minute lag phase (**Figure 17 a** black curve). The lag phase of CsgA polymerization was dramatically shortened when 4% CsgB fibrillar aggregates corresponding to 320 nM of protein were added to the reaction (**Figure 17 a** red curve). This result indicates that the CsgB-wt fibers are capable of seeding soluble CsgA.

9.1.10 CsgB-wt can nucleate CsgA in vitro

CsgB presumably does not form fibrils *in vivo*. Therefore to mimic the in vivo physiological condition for nucleation, soluble and monomeric CsgB instead of CsgB fibers was used to nucleate soluble CsgA. We asked if CsgB-wt monomers provided a nucleating surface for CsgA polymerization. Freshly purified 10μM solution of CsgA was nucleated with different concentrations (100nM,500nM,1000nM) of monomeric freshly purified CsgB-wt which was desalted at pH3.Appropriate amount of soluble monomeric CsgB at pH 3 was mixed with soluble monomeric CsgA and the pH was adjusted to pH 7.4 with the thioflavin T and the aggregation was monitored. The lag phase of CsgA polymerization was dramatically shortened in a concentration dependent manner of the nucleator (**Figure 17b**). This novel finding indicates that monomeric CsgB can be a potent nucleator for CsgA aggregation. This experiment therefore clearly differentiated between the seeding and nucleation processes by CsgB.

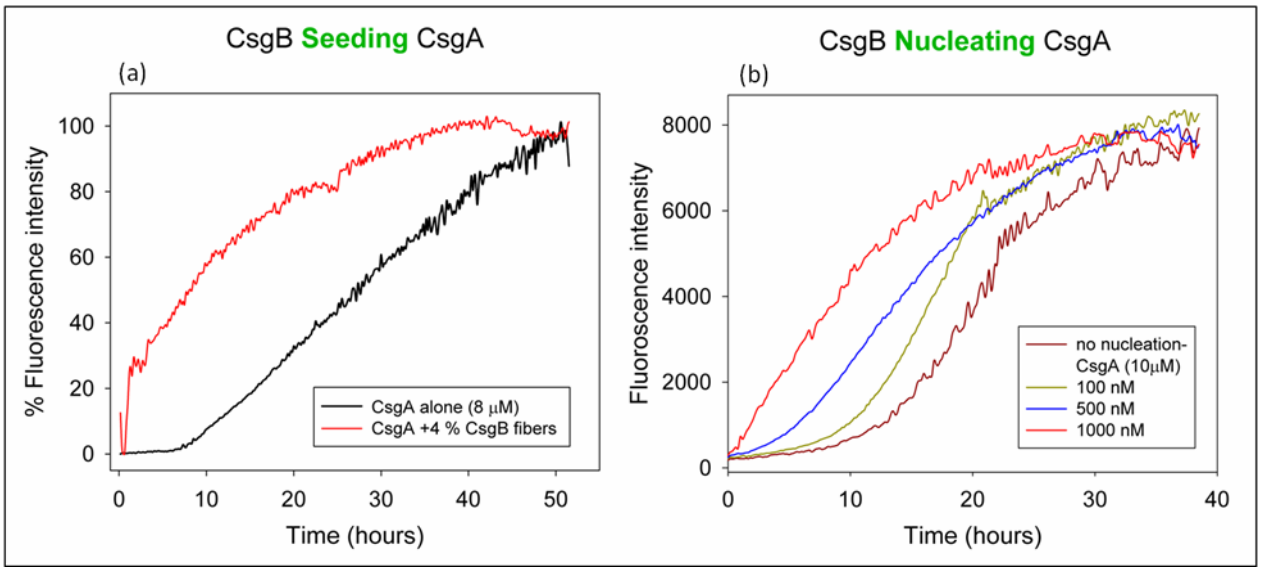


Figure 17: Differentiating CsgB seeding and nucleation of CsgA.

(a) Polymerization of CsgA-wt (8 μM) followed by ThT Thioflavin T (black curve) and 8 μM CsgA in the presence of 4% or 320nM of CsgB-wt fibers (red curve). **(b)** Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T (brown curve) and 10 μM CsgA in the presence of 100, 500, 1000nM of soluble and monomeric CsgB-wt.

9.1.11 Sequence-Specific Analysis of Secondary Structure Elements of CsgB-wt

In order to elucidate the fold of the amyloid fibrils formed by CsgB , protein expression was carried out in minimal medium supplemented with $15NH_4Cl$ and 13C-glucose to obtain CsgB-wt that was 15N and 13C labeled. After cultivation and cell harvest, cells were lysed under native conditions and the insoluble material was solubilized subsequently. The solubilisation was followed by IMAC under denaturing conditions using Ni-Sepharose. Eluted protein was then spun at 50000 rpm for 2 hours at 25° C to remove any preformed seeds and finally desalted using Sephadex G25 gel-filtration column in to the potassium phosphate buffer at pH 3. To ensure complete buffer exchange conductivity was monitored. Only protein fractions devoid of denaturant were pooled. A total concentration corresponding to 10mg of protein was obtained per liter of labeled media. The denatured protein was then desalted into native buffer to remove the denaturant and further fibrillised by increasing the buffer pH to 7.4. To achieve sequence-specific backbone-resonance assignment, 15N and 13C labeled protein fibers were solubilized in a mixture of DMSO and 0.1% TFA enabling complete monomerisation of the fibers. This mixture was then used to record 2D experiment (HSQC). In addition to 3D

experiments HNCA, HN(CO)CA, HNCACB and NOESY . 92 % of CsgB-wt resonances were assigned to the sequence illustrated in **Figure 18**.

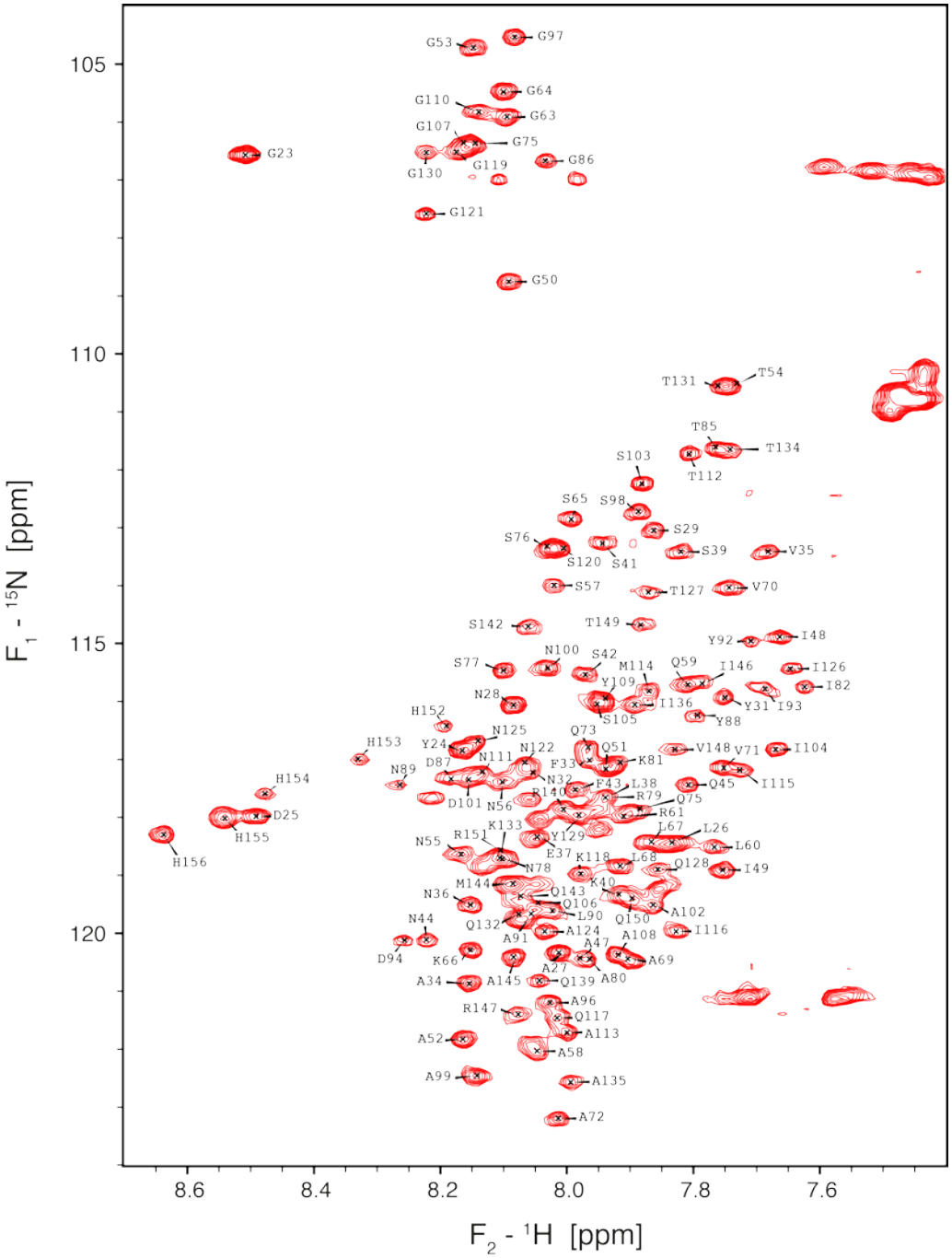


Figure 18: Backbone NH-assignments of CsgB-wt.

ctHSQC spectrum of uniformly 15N labeled CsgB-wt (20 mg wet pellet of fibrils) in d_6-DMSO containing 0.1 % (v/v) d_1-TFA. Sequence-specific chemical shift assignments are indicated next to the corresponding resonance

9.1.12 H/D Exchange NMR Analysis of CsgB Fibrils

Due to the large size of amyloid fibrils, direct structural investigation of these assemblies is very challenging and experimentally difficult. However, the fibrils can be conveniently studied by H/D exchange experiments followed by dissociation of the aggregates and detection of solvent protection patterns in the monomeric state. A concept of such an experiment was introduced in material methods. An important parameter that requires optimization in the quenched H/D exchange technique is the composition of the dissolving buffer, as the desolution of fibrils to monomer is fundamental for the success of the experiment. The ideal buffer should quickly and fully dissociate the fibrils and provide conditions under which exchange in the monomeric stage is minimal. This is typically achieved by mixtures of deuterated dimethyl sulfoxide (DMSO) and 0.1 % trifluroacetic acid (TFA)

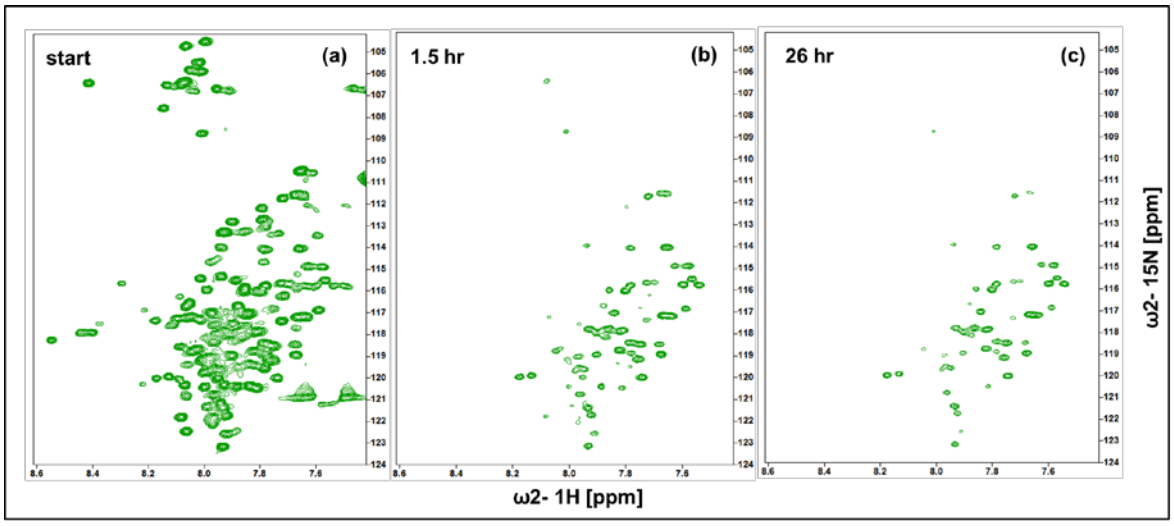


Figure 19: H/D exchange of CsgB-wt.

ctHSQC spectra of uniformly 15N labeled CsgB-wt (~20 mg wet pellet of fibrils) in d_6-DMSO containing 0.1 % (v/v) d_1-TFA. **(a)** The spectra correspond to fully protonated (start) and partially hydrogen exchanged **(b)** 1.5 hours exchange fibrils **(c)** 26 hours exchange fibrils.

Primary results with the HDX revealed most (80%) of the residues being exchanged completely within 24 hours indicating the presence of non specific aggregates instead of ordered fibers. To exclude non specific aggregates the fibrillized samples were washed with 0.1 % SDS, since the ordered fibers are resistant to the SDS wash only the non specific aggregates were dissolved by SDS. For further HDX experiments the protein

samples were fibrillized at a very low concentration corresponding to approximately 30 µM and washed with SDS to enrich the ordered aggregates.

Quenched hydrogen-exchange allows the identification of solvent-protected backbone amide protons indicative of hydrogen bond formation. **Figure 19a** shows the reference [15N, 1H] correlation NMR spectrum of CsgB in DMSO corresponding to fully protonated CsgB fibrils. The spectrum contains one cross-peak for each backbone 15N-1H moiety of CsgB, enabling a residue-specific determination of the hydrogen exchange rates. After 1.5 hour of H/D exchange with the CsgB fibrils (potassium phosphate, pH 7.4) approximately 40% of the resonances are significantly reduced or absent from the spectrum (**Figure 19b**) and after 26 hours 70% resonances are significantly reduced or absent from the spectrum (**Figure 19c**). The corresponding amides have exchanged with solvent deuterons that are not observable in the [15N, 1H] correlation NMR experiment. CsgB fibrils revealed protonation state of 60% of the residue after 24 hours indicating its slow exchange rate. This observation is consistent with the qualitative notion that CsgB fibrils contain a highly protected core structure (14, 15).

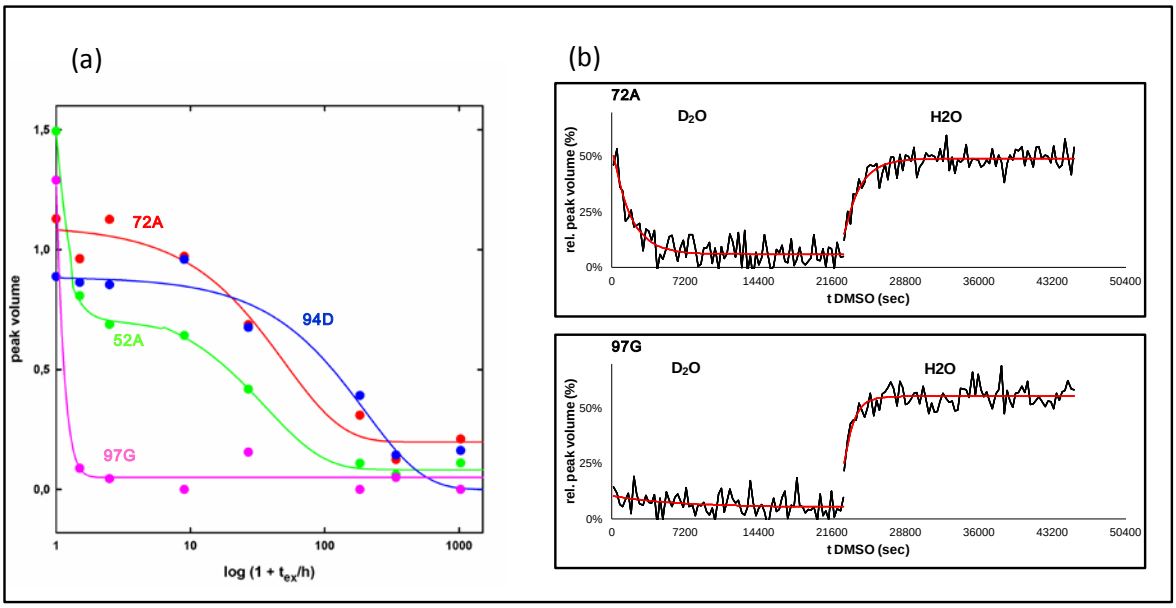


Figure 20: Determination of H/D exchange rates of CsgB-wt.

(a) Monoexponential and Biphasic fit of peak volumes at t_{DMSO} = 0 min to t_{ex} corresponding to selected backbone amides. The biphasic fit shows sum of two monoexponential equations **(b)** Intrinsic H/D exchange in DMSO of 72A (upper panel) and 97G (lower panel) backbone amides. The evolution of peak volumes after monomerising the CsgB-wt fibrils in DMSO/TFA and final addition of H_2O is shown. Red curves show the monoexponential fits of the raw data.

40 % of the residues showed fast exchange with deuterium hence disappeared within 24 hours. [15N-1H]-correlation NMR spectra of CsgB in DMSO corresponding to the remaining fully protonated and partially H/D-exchanged CsgB fibrils was further used to determine the accurate peak intensity of the backbone amides. 15N labeled CsgB fibrils chosen for H/D exchange NMR analysis were incubated in D2O buffer for various time points up to 20 weeks (0 h, 30 min, 1.5 h, 8 h, 24 h, 1w, 2 w, 6 w, 20 w). The fibrils incubated at each time point were dissolved in DMSO followed by recording series of 80 spectra over a time frame of 6 hours (**H/D Exchange NMR Experiments**). The rate of decrease in peak intensity for each residue after dissolution in DMSO due to residual D_2O was determined and extrapolated to get accurate peak intensity for that corresponding time point (**Figure 20b** upper panel). Further, addition of H_2O followed by recording the spectra helped in determining the residues which showed faster exchange rate in DMSO (**Figure 20b** lower panel). Hence, plotting the accurate peak intensity determined by series of [15N-1H]-correlation spectra for each time point of each residue against various times point gave the exchange rate of each residues in CsgB fibrils (**Figure 20a**).

55% of CsgB residues displayed biphasic behavior in the exchange plot (**Figure 20a** for example amino acid 52A) indicating heterogeneity (Luhrs et al., 2005; Vilar et al., 2008). 45% of CsgB residues displayed a monoexponential decay (**Figure 20a** for example amino acid 72A, 94D & 97G) suggesting a well defined and homogeneous structure of the fibrils. The residue showing biphasic behavior indicates presence of two different types of conformation. The backbone amides showing biphasic behavior constituted a subpopulation having fast H/D exchange as well as the second subpopulation displaying slow exchange rate (**Figure 21a &c**). Exchange rate of both populations was analyzed separately (**Figure 21b & d**). To identify the major population below equation was used.

$$\text{Major population} = \frac{I_{0\ \text{FAST}} - I_{0\ \text{SLOW}}}{I_0} * 100$$

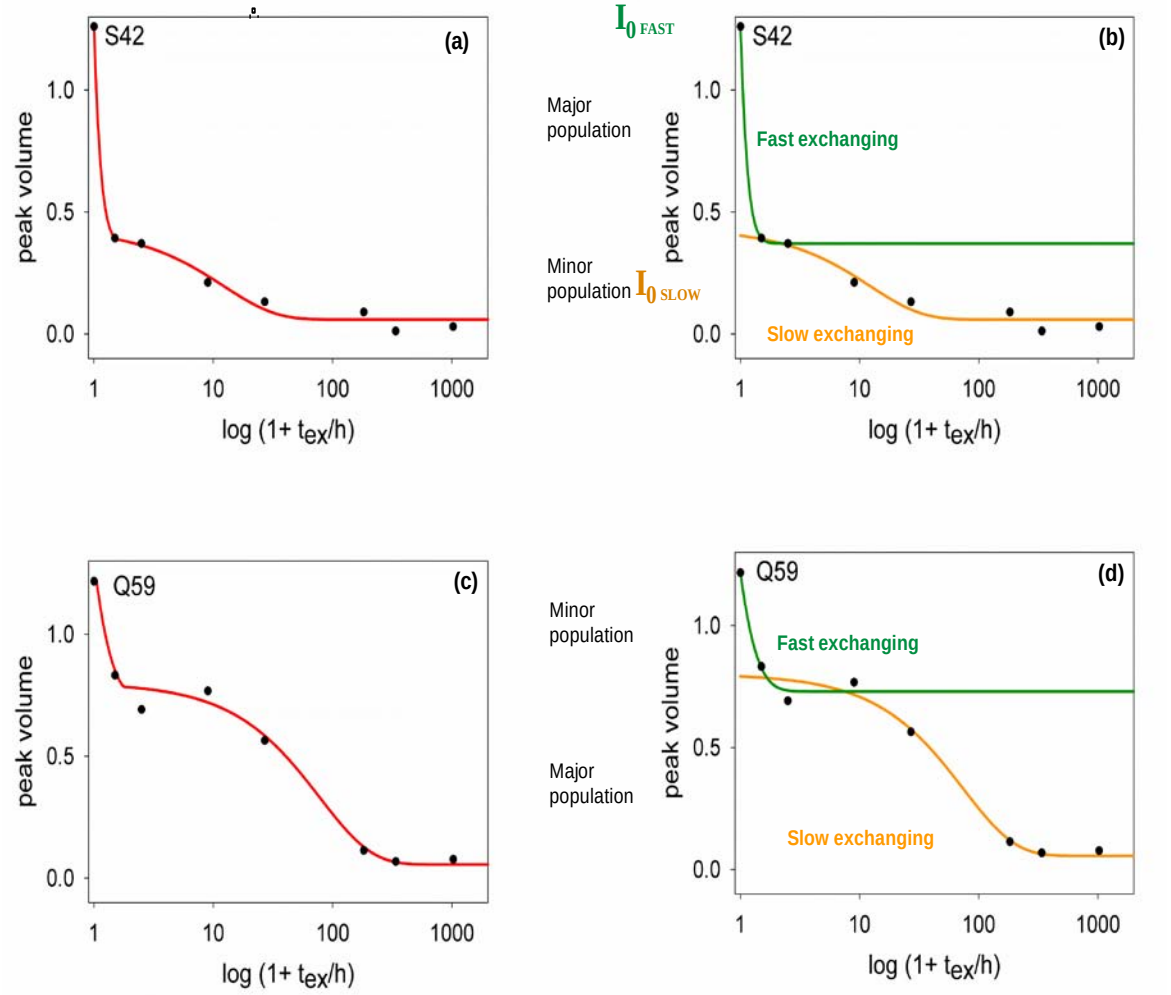


Figure 21: Determination of different populations in heterogenous fibers of CsgB-wt

(a) Peak volumes displaying biphasic fit of corresponding to amino acid residue S42 shows fast exchanging major population and slow exchanging minor population **(b)** Two monoexponential rates calculated from the biphasic fit for both major and minor population of S42. Both $I_{0\ FAST}$ & $I_{0\ SLOW}$ are shown in the picture **(c)** Peak volumes displaying biphasic fit of corresponding to amino acid residue Q59 shows fast exchanging minor population and slow exchanging major population **(d)** Two monoexponential rates calculated from the biphasic fit for both major and minor population of Q59.

The exchange rate for both the subpopulations of CsgB was calculated. This helped in determining the I_0 or the initial intensity (peak volume) of the two different subpopulations. The difference between the peak intensity of the subpopulation divided by the peak intensity at zero time point give the percentage of slow exchanging population. Any subpopulation displaying more than 50% normalized intensity was termed as major and the other as minor. The relative subpopulation for each residue exhibiting biphasic behavior was analyzed, normalized and plotted. The major population was shown in green while the minor population was shown in grey (**Figure 23**). Hence the exchange rate of the major population was shown in green while the minor population was shown in grey (**Figure 22**).

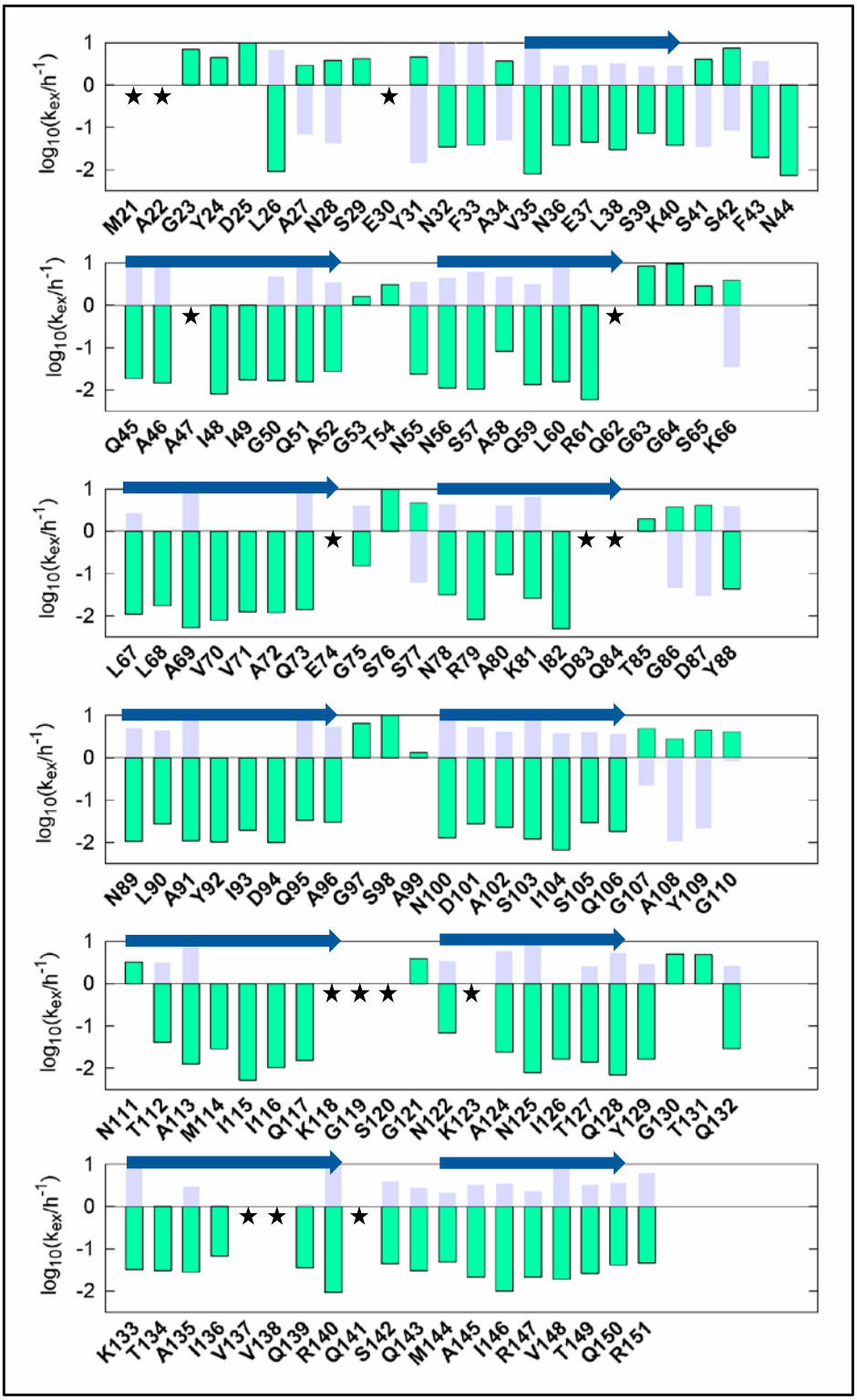


Figure 22: H/D exchange NMR data of CsgB-wt and secondary structure prediction.

Green and grey bars indicate exchange rates of CsgB-wt k_{ex} (h^{-1}). Grey bars indicate exchange rates of the minor population and green indicate major population in case of biphasic behaviour. Residues marked with a star are missing assignments. Blue arrows indicate β-strands.

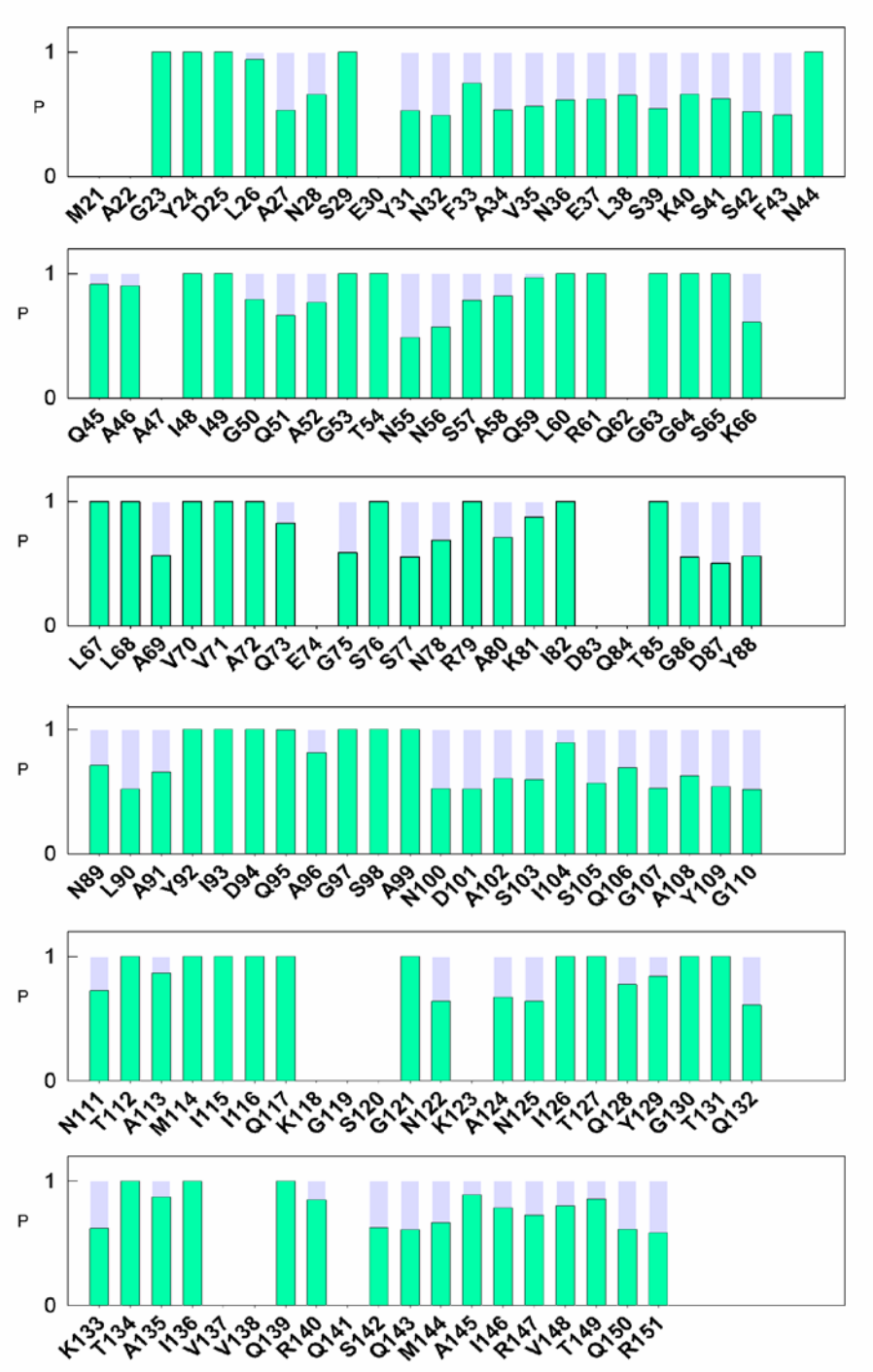


Figure 23 : Relative populations in heterogeneous fibers of CsgB-wt

Grey bars indicate the minor population and green indicate major population in case of biphasic behaviour. Residues marked with a star are missing assignments.

Residues for example I49, V71, A72, and D94 showed monoexponential decay with slower exchange rates (**Figure 22**) indicating uniform incorporation between the different conformations that existed inside the amyloid core. While other residue for example G97, T54,

S98, Y24 and N111 showed monoexponential decay with faster exchange rates (**Figure 22**) indicating its presence outside the amyloid core possibly in the loop region of the amyloid structure. However majority of the residue like L38, Q51, S57, A69 and A91 showed biphasic behavior indicating heterogeneity (**Figure 22**). The heterogeneity in the residue implies that it's been incorporated in the fibers in a different way between different conformations or incorporated differently in a sub fraction of fibrils forming varied hydrogen bonds within. Interestingly the exchange rate of the R5 repeat of CsgB showed a complete heterogeneous behavior indicating varied pattern of hydrogen bond formation between various conformations of fibrils.

The level of heterogeneity was lower near the predicted loop region. Some of the residues include L26, N32 and F33 showed extreme overlapping with the others slow exchanging residues like L67, N122 and Q51 making impossible to analyze accurate exchange rate (**Figure 18**). Collectively the exchange rate of the major subpopulation revealed secondary structure of CsgB consisting of two β-strands that are connected by a short β-turn for every repeating units from R1 to R4 (**Figure 22**). The β-strands of adjacent repeats are connected by flexible loops comprising four to five residues. Amyloid core within repeating units R1 to R4 showed 4 pairs of beta strands, the fifth repeat shows a complete heterogeneous behavior which can be implicated with the lack of conserved sequence alignment (**Figure 10**) and high percentage of positively charged amino acid which may be involved either in binding to the cell membrane or interacting with CsgF. The secondary structure of CsgA consists of N-terminal flexible region followed by two β-strands that are connected by a short β-turn for every repeating units from R1 to R5 (unpublished data by Agnes Zimmer). On contrary a novel finding showed that CsgB consists of an additional beta strand in the N-terminal region (**Figure 22**). The sequence specific secondary structure revealed the presence of four pairs of beta strand within the fiber core comprising of residues 45-132 which fulfill specific roles (**General Amyloid features**) in directing amyloid formation. Interestingly, the N-terminal flexible region 21-44 showed an additional single beta strand and the C-terminal 133-151 showed complete protection with higher level of heterogeneity. These results show a difference in the N and C terminal secondary structure of CsgB contrary to the predicted model (White et al., 2001). This novel finding showing varied N and C terminal structures of CsgB-wt prompted us to investigate their influence on the structure and kinetic properties further in the following chapters.

9.1.13 Discussion

Evidence suggests that in case of disease related amyloids, formation of the amyloid precursors during the amyloidogensis are cytotoxic rather than the fibrillar end product of the misfolding event (Bucciantini et al., 2002; Larson et al., 1999; Lue et al., 1999; Malisauskas et al., 2005; McLean et al., 1999; Moechars et al., 1999; Sirangelo et al., 2004; Sousa et al., 2001; Wang et al., 1999). In order to overcome the above mentioned risk functional amyloid biogenesis has evolved specialized mechanisms to eliminate the build up the cytotoxic intermediates. For example, in a mammalian functional amyloid PMEL17, it is hypothesized that the aggregation of the monomeric proteins is facilitated so rapidly that no toxic intermediates are formed (Fowler et al., 2006). Unlike functional amyloid, curli has evolved a unique amyloid pathway comprising of a dedicated nucleator protein in vivo. Having a dedicated nucleator may be an alternative strategy to decrease the formation of cytotoxic intermediates by providing the uniform template for fiber formation .Nucleus formation is proposed to be the rate limiting step during amyloid fiber formation (Rochet and Lansbury, 2000b).

In the curli biogenesis in E.coli, CsgA is secreted from the cell as a soluble, unpolymerized protein. CsgA would remain unpolymerized unless it contacts the nucleator protein CsgB which mediates the conversion of CsgA from a soluble protein to an insoluble fiber. Once fiber formation is initiated, the growing fiber tip acts as a template for soluble monomeric CsgA for additional polymerization in a process called elongation. This part of the results confirmed the amyliodogenic properties of full length CsgB-wt including beta sheet rich secondary structure, binding to the thioflavin T an amyloid biding dye. The *in vitro* aggregation kinetics CsgB was faster compared to CsgA corroborating its role as a nucleator. The enhanced folding property of CsgB results in faster aggregation and vice versa, hence provides the right kind of template and avoiding the non specific random seed formation during fiber growth *in vivo*. This process may help in preventing formation of heterogeneous fibers and possibly toxic intermediates.

The protein sequence of CsgB shows high number of charged amino acid of total 19 out of which 11 are positively charged amino acid compared to CsgA which shows a total of 16 charged amino acid out of which 4 are positively charged. This higher percentage of positively charged amino acid in CsgB especially in the C-terminal region of R5 repeat makes it highly positively charged compared to CsgA. The CsgB-wt also displays a calculated pI of 9.30 compared to CsgA which has calculated pI of 5.37 this difference in pI would play a major role in protein solubility at pH 7 and at pH 3. Since the curli protein including CsgB and CsgA start as random coil calculated pI would play a major role in folding of protein. Higher percentage

of charged amino acid in CsgB indicates that ionic interaction would play a major role in the aggregation process of CsgB. The aggregation monitored under low pH showed a major role of pH in the aggregation process of CsgB contrary to CsgA (**Figure 16**) (Dueholm et al., 2011). The above evidence concludes that the aggregation and/or folding of CsgB is predominantly driven by ionic interaction leading to faster folding of CsgB which is highly positively charged, which would then nucleate the highly negatively charged CsgA on the cell surface.

CsgA polymerization in vitro contains a 200-300 mins lag phase that can be shortened by either addition of preformed CsgA or CsgB fibers. A novel finding in this part of the result showed that CsgA polymerization was shortened by soluble monomeric CsgB-wt which acts as nucleator. The above finding led to furnish a refined model of extracellular nucleation/precipitation where both CsgA and CsgB are secreted to the cell surface. CsgB folds faster compared to CsgA and nucleate CsgA to form fibers. During which single molecules of CsgB are incorporated in between the long oligomers of CsgA which was shown by immunolabelling methods (Bian and Normark, 1997). The results obtained clearly demonstrated that monomeric and soluble CsgB act as nucleator of CsgA during in vivo curli biogenesis, contrary to the previous belief that CsgB would form oligomers or fibers on the cell surface which would later provide the platform for CsgA to form fibers (Hartl et al., 2011).

The secondary structure analysis of the CsgB fibers by HDX NMR showed higher level of heterogeneity. This high level of heterogeneity in the fiber structure could be of two main reasons

1- *In vivo* the folding of CsgB is influenced by the presence of charged surface on the cell membrane and also the other curli biogenesis proteins including CsgE, CsgF and CsgG.

2- Due to the robust and fast self folding propensity of CsgB, it may lead to secondary nucleation resulting in formation of heterogeneous sample at high concentration *in vitro*.

3- Formation of CsgB fibers could be also artificial when compared to the in vivo condition, which has been corroborated with the previous results where CsgB was able to nucleate CsgA in a monomeric form

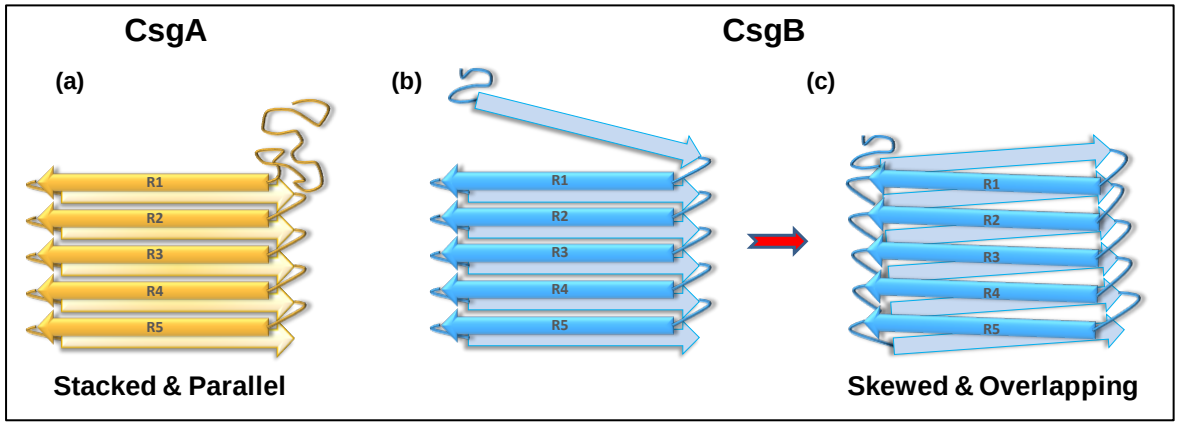


Figure 24: Suggested model for the orientation of the beta strand in CsgA and CsgB-wt.

(a) Predicted model of stacked and parallel arrangement of beta strand in CsgA based on NMR HDX data **(b)** Presence of extra beta strand in CsgB-wt based on NMR HDX data **(c)** Predicted model of skewed and overlapping arrangement of beta strand in CsgB-wt based on NMR HDX data.

The secondary structure of full length CsgB revealed unique N and C terminal on contrary to the predicted model which suggested the structure of CsgA and CsgB to be similar (Collinson et al., 1999b; White et al., 2001).

N-terminal flexible region (21-44) of CsgB-wt showed a single beta strand in addition to the C-terminal (133-151) region which shows complete protection with higher level of heterogeneity due to high percentage of charged residue. Based on these results we could hypothesize a nucleation model where CsgB showing overlapping or skewed (neither parallel nor intersecting) organization of beta sheet elements rather than stacked (**Figure 25**). The HDX data the secondary structure of full length CsgA (unpublished data Agnes Zimmer) revealed five pairs of beta strands forming five distinct repeats followed by highly flexible N-terminal. Each repeat forms a segment of two β-strands, which are connected by a short β-turn. The HDX data of CsgA complements the predicted beta helix model (Collinson et al., 1999b) showing five beta strands stacked and parallel to one another in each repeat (**Figure 24 a**).

The secondary structure of full length CsgB revealed similar amyloid core as CsgA comprising of five pairs of beta strands forming five distinct repeat , but on contrary to the predicted model (White et al., 2001) CsgB showed an extra beta strand in the N-terminal region as revealed by HDX data. This extra beta strand makes CsgB to form five and a half repeat (**Figure 24 b**) rather than five repeats, this leads to the skewed (neither parallel nor intersecting) and

overlapping arrangement of the beta strand (**Figure 24 c**). The extra beta strand flips providing a sequence selective interface for the nucleation of unfolded soluble CsgA. This arrangement helps not only in specific nucleation of CsgA but also gives directionality for the fiber growth (**Figure 25 a**). The above hypothesis is further proved by using the truncation mutants, which forms the basis of next chapter.

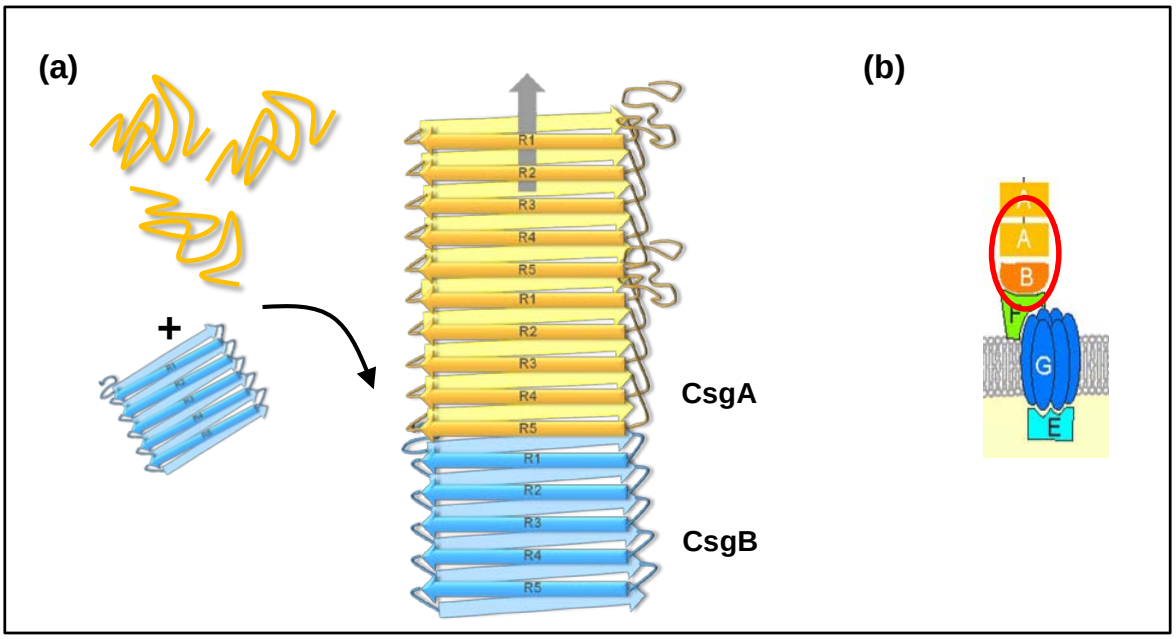


Figure 25 : Suggested model for nucleation of CsgA by CsgB.

(a) Soluble and random coil CsgA (yellow coils) nucleated by folded and monomeric CsgB-wt (blue colour) .Predicted model of skewed and overlapping arrangement of beta strand in CsgB-wt nucleating CsgA by providing sequence selective interface for the nucleation **(b)** Schematic representation of CsgB-wt nucleation on the cell surface.

9.2 Chapter 2- N and C-terminal repeating units govern CsgB aggregation and Nucleation responsiveness

9.2.1 Abstract:

Nucleation plays an important role in curli biogenesis; therefore structure of CsgB-wt may yield a detailed structure function relationship of nucleation process. Secondary structure analysis of CsgB-wt by quenched HD exchange revealed distinct N and C terminal regions. Deletion analysis was used to elucidate the importance of N and C terminal regions in the process of nucleation. Deletion analysis of 20 amino acids residues from N-terminal region (21-40) that include beta strand, showed enhanced nucleation capacity compared to full length CsgB-wt. The truncation involving 20 amino acid residues from C-terminal region (132-151) showed decreased nucleation capacity compared to full length CsgB-wt. The truncation analysis of ΔN & ΔR5 CsgB-truncation mutants revealed that the N- terminal of CsgB was involved in sequence selective interface for nucleating CsgA and C-terminal was involved in the self folding propensity of CsgB which later proves to be the major rate limiting step during the nucleation of CsgA. Biophysical investigation of the deletion mutants of CsgB by CD, FTIR showed the secondary structure rich in cross beta sheet and displayed similar amyloidogenic properties as compared to the CsgB-wt. Even though the deletion mutants showed varied nucleation capacity, the sequence specific secondary structure of the fibers showed similar amyloid core as revealed by HD exchange. The deletion analysis further proved the hypothesis that CsgB had skewed (neither parallel nor intersecting) and overlapping organization of beta sheet elements providing a sequence selective interface for the nucleation of CsgA.

9.2.2 Starting Point of CsgB-truncation analysis

The amino acid sequence alignment of CsgA and CsgB shows varied conserved sequences indicating different structural assembly. Especially both N and C terminal region of CsgA and CsgB were conserved differently (**Figure 10**).

Based on the secondary structure analysis by HDX NMR of full length CsgB-wt, both N and C terminal domain were distinguished for the truncation analysis.

9.2.3 CsgB ΔN & ΔR5 truncation mutants protein purification

The truncation constructs both N-terminal (ΔN) and C-terminal (ΔR5) CsgB were cloned in a pet11d vector and was affinity purified under denaturing conditions to obtain pure and monomeric CsgB truncation mutants with approximate yields of 8 mg per liter of LB as described in material & methods. The protein was desalted in to native buffer at low pH to get a monomeric soluble truncated CsgB.

9.2.4 Production of monomeric soluble CsgB-truncation mutants using pH switch

Like full length CsgB, the aggregation of truncated CsgB mutants at pH3 was drastically reduced (**Figure 28**) indicating that self folding and/or aggregation of CsgB depend mainly on ionic interaction between the residues. The aggregation of CsgB was faster at higher pH (**Figure 28**) showing the amyloidgenic property can be controlled using pH as switch even for the truncation mutants.

9.2.5 CsgB truncation mutants form amyloid-like fibers *in vitro*

We focused on characterizing the biochemical and biophysical properties of CsgB truncation mutants to compare with the CgsB-wt for assessing the effects of mutation. Mutants were cloned into the expression vector pet11d and were affinity purified from cell free supernatants as described (material &methods). Immediately after desalting to pH 3, CsgB-truncation mutants were SDS soluble and therefore migrate to its predicted molecular weight in an SDS-PAGE gel (Fig. 26 a).

The denaturant was removed by desalting using sephadex G25 gel filtration column in to 5 mM potassium phosphate buffer at pH 3, immediately followed by changing the buffer conditions to 50mM potassium phosphate pH7.4, using the 1M stock of potassium phosphate pH 7.4. Increasing the pH and incubation at room temperature showed the formation of precipitate instantly; this was then allowed to aggregate for 24 hours. Interestingly CsgB ΔN mutant showed visible precipitate instantly after increasing pH similar to CsgB-wt. In contrast CsgB ΔR5 mutant showed a delayed precipitation

indicating the slower aggregation property of CsgB ΔR5 mutant. These precipitates were later analyzed by biophysical methods to characterize and to compare the amyloidogenic properties of the truncation mutants with full length CsgB-wt (21-151).

9.2.6 Morphology of the precipitates

The ultrastructure of aggregates from truncation mutants aggregates was investigated by transmission electron microscopy (TEM). Freshly purified CsgB truncation mutants were desalted in pH 3 buffer which was then adjusted to pH 7.4 using 1M potassium phosphate stock solution. TEM analysis of purified 30 µM CsgB truncation mutants revealed the presence of highly ordered, amyloid-like fibers similar to the CsgB-wt (**Figure 26 b, c**).

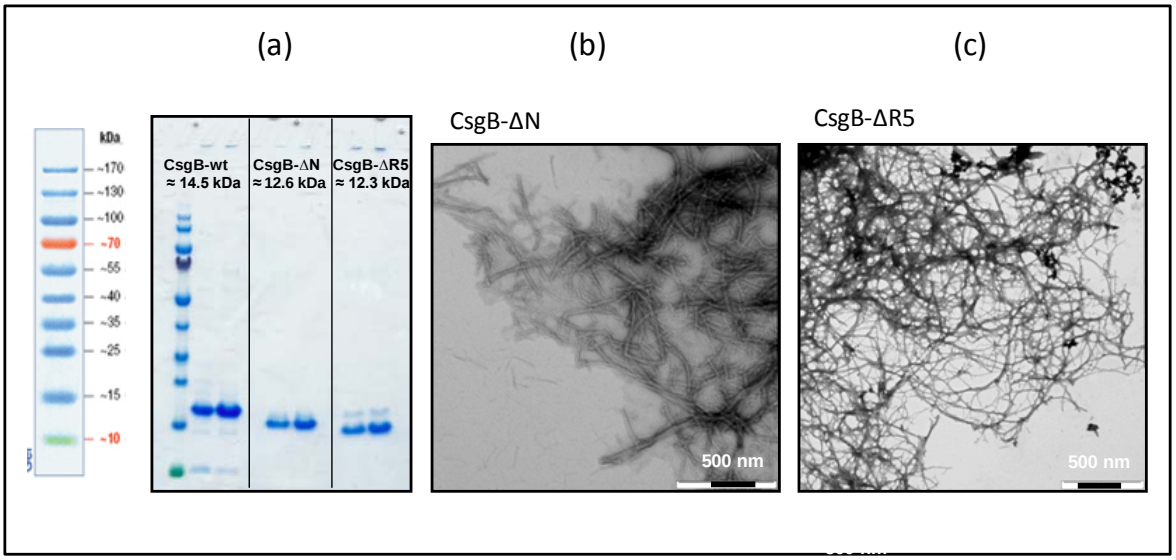


Figure 26 : Biochemical property of CsgB-truncation mutants.

(a) Coomassie stain SDS-PAGE of CsgB-wt (14.5 kDa), CsgB-ΔN (12.6 kDa) and CsgB-ΔR5 (12.3 kDa) samples. Molecular size markers (in kilo daltons) are indicated on the left. **(b)** Negative-stain EM micrograph of purified CsgB-ΔN **(c)** Negative-stain EM micrograph of purified CsgB-ΔR5. Magnification indicated on the right corner (500nm)

9.2.7 Secondary structure analysis CsgB-truncation mutants

The structural changes that occur during the polymerization of the truncation mutants were measured using Fourier transform infrared spectroscopy (FTIR). Freshly purified truncated CsgB was desalted in buffer with pH 3 which was then adjusted to pH 7.4 using 1M potassium phosphate stock solution. The secondary structure of the protein was determined using BioATR–FTIR. The FTIR spectra showed amide I band between 1700 cm^{-1} and

1600 cm−1. Following incubation of CsgB-truncation mutants at room temperature at pH 7.4 for 24 hours the maximum of the amide I shifted to 1620 cm−1 with an additional low - intensity peak at 1659 cm−1 (**Figure 27**). The FTIR spectrum of both the CsgB-trunc ΔN (**Figure 27** green curve) and CsgB-trunc ΔR5 mutant (**Figure 27** red curve) was similar to full length CsgB-wt (**Figure 15 a**) indicating similar secondary structure.

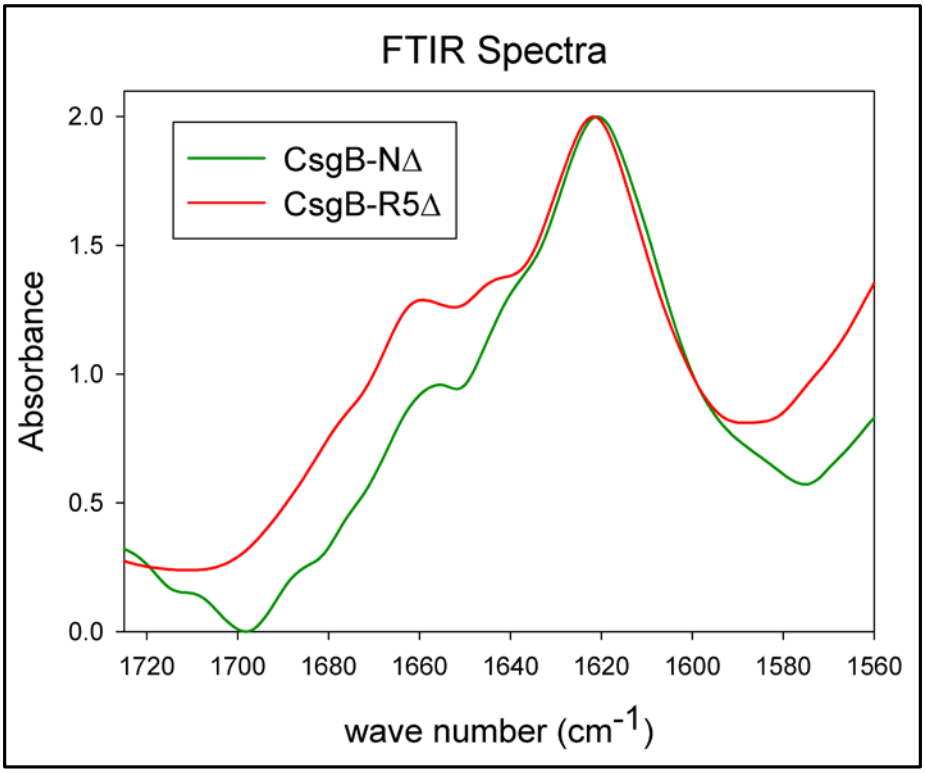


Figure 27 : Biophysical characterization of CsgB-truncation mutants.

FT-IR spectra showing the amide I band of aggregates (1 mg/ml) formed by CsgB-truncation mutant samples. CsgB-ΔN (green curve) and CsgB-ΔR5 (red curve).

9.2.8 CsgB-truncation mutant's aggregation kinetics

Both CsgB-trunc ΔN and ΔR5 was used to check its binding with the amyloid-specific dye, thioflavin T and to monitor its aggregation kinetics through the increased fluorescence upon dye binding. Freshly purified CsgB-truncation mutants in varied concentration (1, 2, 3,5,7,9 μM) were mixed with thioflavin T simultaneously increasing the pH to 7.4. The thioflavin T biding showed increased fluorescence in a concentration dependent manner. The aggregation kinetics of CsgB-trunc ΔN (**Figure 28c**) appeared to be almost similar compared to the CsgB-wt (**Figure 28a**) with very less or nearly no lag phase. In contrast the aggregation kinetics of CsgB-trunc ΔR5 (**Figure 28e**) showed a much delayed lag phase of about 200-300 mins, indicating slower aggregation profile. Similar to the wild

type CsgB-wt (**Figure 28b**) both the CsgB-truncation mutants ΔN (**Figure 28d**) and ΔR5 (**Figure 28f**) showed no or little aggregation at pH3.

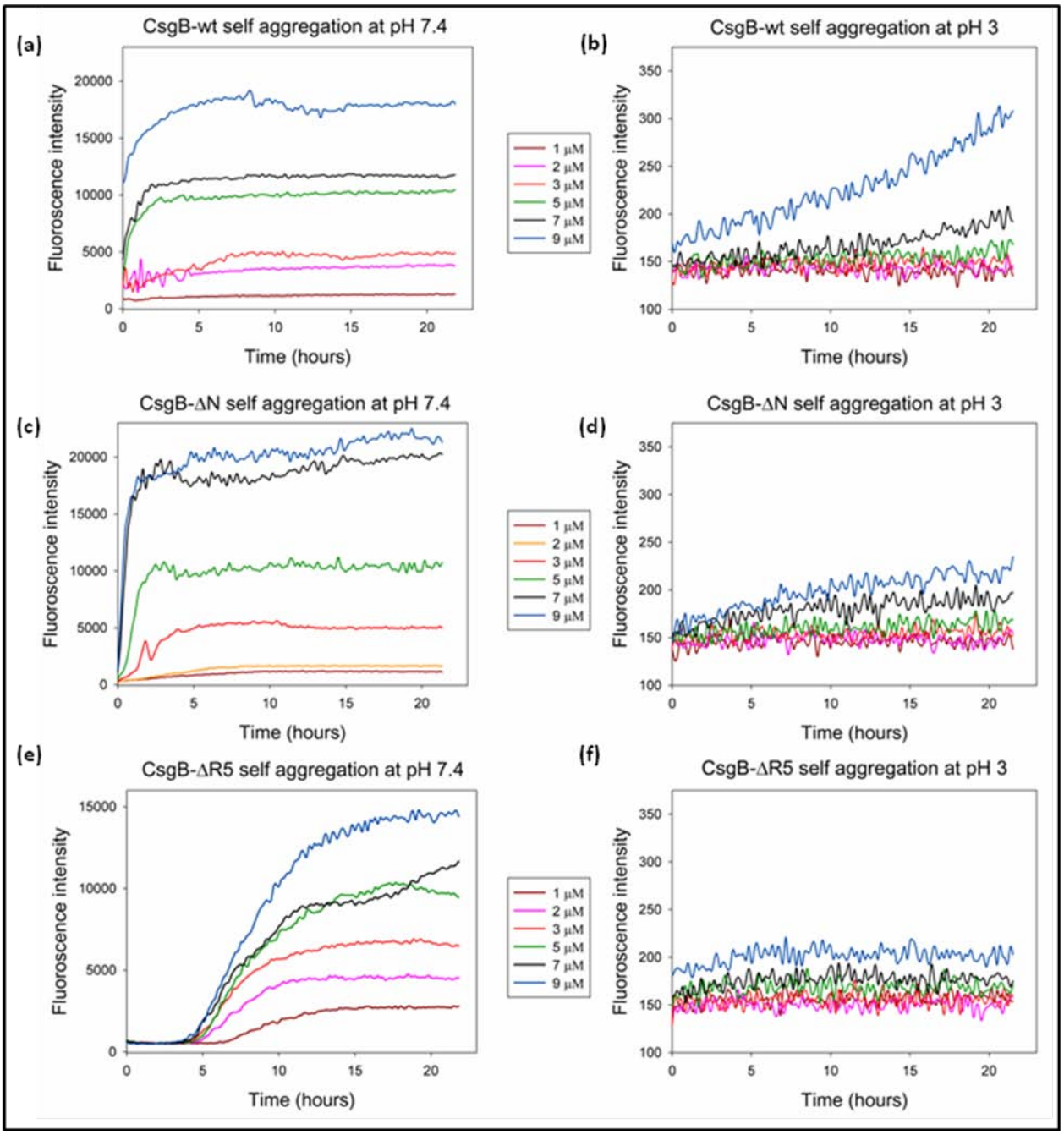


Figure 28: Aggregation kinetics monitored by Thioflavin T binding.

(a) (c) (e) Fibrillation of CsgB-wt, CsgB-ΔN and CsgB-ΔR5 followed by ThT at pH 7.4. Increase in ThT fluorescence depending on time and protein concentration (1, 2, 3,5,7,9 μM). CsgB-ΔR5 shows delayed aggregation with the lag phase **(b) (d) (f)** Fibrillation of CsgB-wt, CsgB-ΔN and CsgB-ΔR5 followed by ThT at pH 3. Little or no increase in ThT fluorescence depending on time and protein concentration (1, 2, 3,5,7,9 μM).The 'X' is reduced according to the Thioflavin T fluorescence to show the expanded curves.

Both CsgB-trunc ΔN and ΔR5 aggregates are β-sheet-rich fibers that interact with the amyloid specific dyes thioflavin T. CsgB ΔR5 truncation mutant shows a delayed aggregation compared to the CsgB-wt and CsgB-trunc ΔN. These biochemical and biophysical features of both the CsgB-trunc ΔN and ΔR5 mutants displayed similar secondary structure as CsgB-wt.

9.2.9 CsgB-truncation mutants can nucleate CsgA *in vitro*

Varied self folding property of CsgB ΔR5 truncation mutant led us to investigate its role in nucleation of CsgA. Freshly purified 10μM solution of CsgA was nucleated with different concentrations (25nM, 50nM, 100nM, 200nM, 500nM, 1000nM) of freshly purified monomeric CsgB-wt, CsgB-trunc ΔN and ΔR5 mutants. The lag phase of CsgA polymerization was dramatically shortened in a concentration dependent manner of the nucleator (**Figure 29**). The nucleation assay revealed that CsgB ΔN (**Figure 29b**) was a better nucleator compared to CsgB-wt (**Figure 29a**) and the truncation mutant ΔR5 (**Figure 29c**) showed the least nucleation capability. To accurately determine the varying nucleation capacities of the mutants, lag time for CsgA polymerization in presence and absence of nucleator was determined (Wang et al., 2008) and plotted against the nucleator concentration of CsgB-wt, CsgB-trunc ΔN and ΔR5 mutants (**Figure 29d**). These results show that 100nM to 200nM of the CsgB-trunc ΔN was enough to reduce the lag phase of CsgA polymerization considerably compared to CsgB-wt and ΔR5 mutant. Nucleator concentration of 500nM and 1000nM showed a clear result indicating that CsgB ΔN showed the highest nucleation capacity followed by CsgB-wt and truncation mutant ΔR5 showed the least nucleation capability (**Figure 29d**).This novel finding indicates that N-terminal region of CsgB may be involved in the process of giving directionality and sequence selective interface during the curli nucleation, further supporting the nucleation model hypothesized based on the secondary structure of CsgB (**Figure 25**).

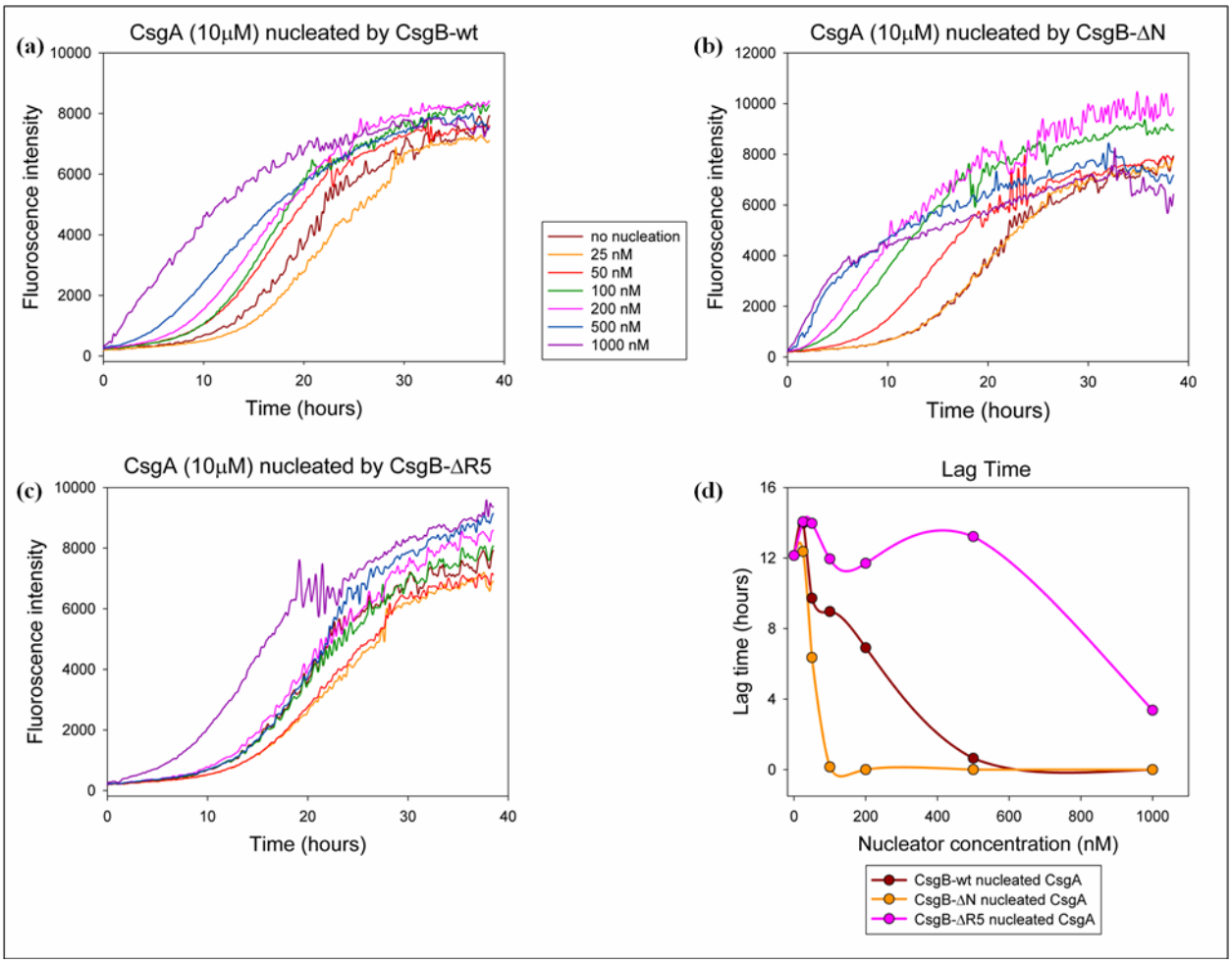


Figure 29: Nucleation of CsgA by CsgB-wt and CsgB-wt truncation mutants.

(a) Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in presence of increasing concentration of monomeric and soluble nucleator CsgB-wt. CsgA without any nucleator used as control **(b)** Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in presence of increasing concentration of monomeric and soluble nucleator CsgB-ΔN **(c)** Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in presence of increasing concentration of monomeric and soluble nucleator CsgB- ΔR5 **(d)** Lag time was plotted against the nucleator concentration for determination of the nucleation capacity of CsgB-wt & mutants.

9.2.10 Nucleation by CsgB-truncation mutants monitored by transmission electron microscopy

To further elucidate the varied nucleation capacity of the CsgB-truncation mutants, the nucleation process was monitored by transmission electron microscopy (**Figure 30**) in combination with Thioflavin T binding (**Figure 31**) over the period of 24 hours. Freshly purified 10μM solution of CsgA was nucleated with 200 nM concentrations of monomeric freshly purified CsgB-wt, CsgB ΔN and ΔR5. The nucleated sample was kept on the shaker at 25°C for polymerization followed by aliquots drawn every 90 mins for over 24hours.

The polymerization was stopped by adding 1 % glutaraldehyde, which was then stored at 4°C. The analysis of each time point through EM revealed a novel finding in the process of curli polymerization. The polymerization begins with the soluble protein being converted to the amorphous aggregates of varying size followed by formation of ordered fibers. The growth of the fibers occurred by the attachment of the amorphous aggregates to the fibers followed by the complete conversion of the amorphous aggregates to fibers within 24 hours.

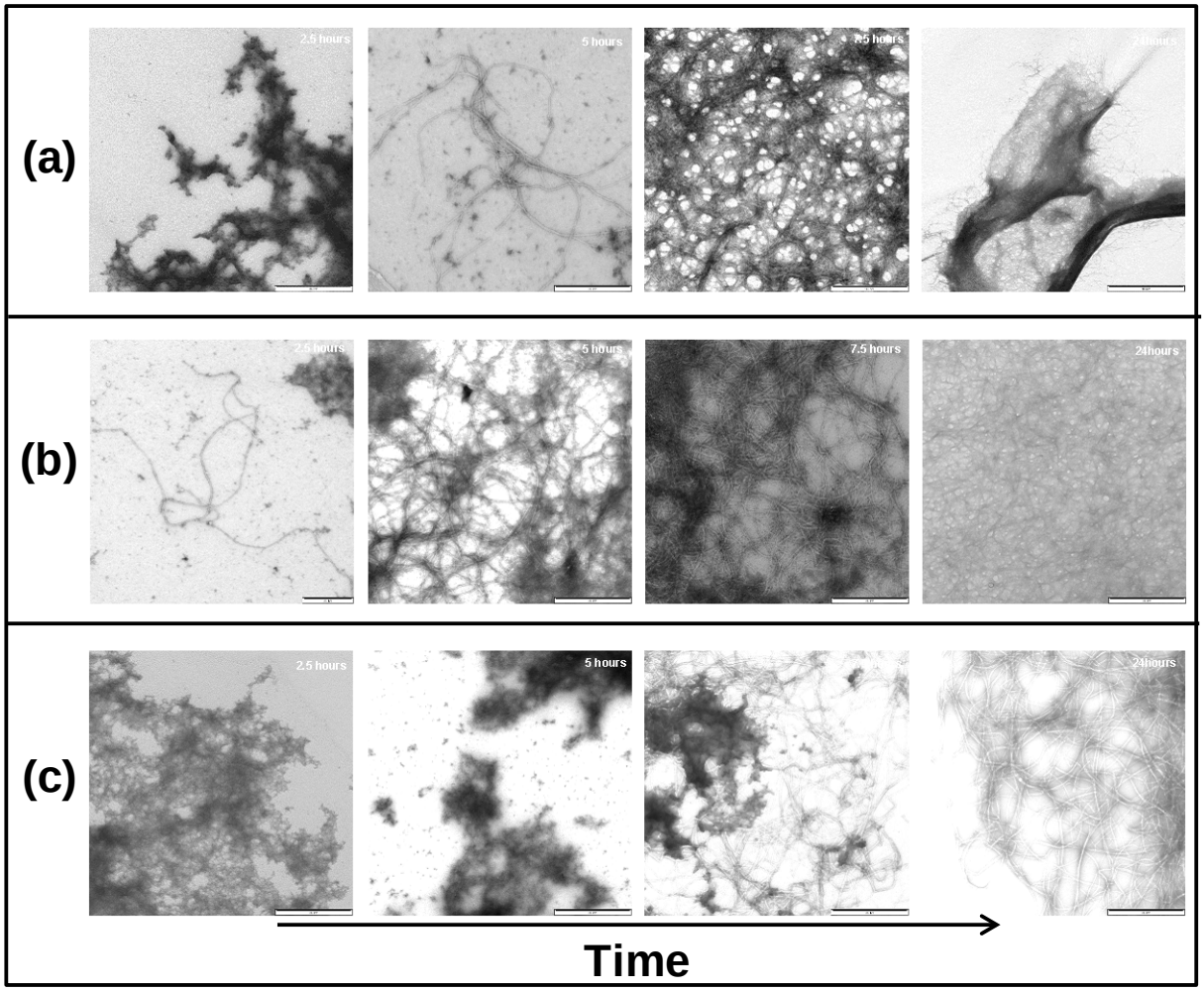


Figure 30: Nucleation of CsgA by CsgB-wt and CsgB-wt truncation mutants monitored by Electron microscopy.

(a) Polymerization of CsgA-wt (10 μM) in presence of 200nM monomeric and soluble nucleator CsgB-wt **(b)** Polymerization of CsgA-wt (10 μM) in presence of 200nM monomeric and soluble nucleator CsgB-ΔN **(c)** Polymerization of CsgA-wt (10 μM) in presence of 200nM monomeric and soluble nucleator CsgB- ΔR5.

The EM analysis of the nucleation process corroborated the nucleation assay monitored by Thio T binding assay indicating that CsgB ΔN showed the highest nucleation capacity (**Figure 30** panel b) which showed fibers within 90 mins followed by CsgB-wt (**Figure 30** panel a) which showed fibers at 180 mins and truncation mutantΔR5 showed the least nucleation capability (**Figure 30** panel c) which showed fibers at 270 mins.

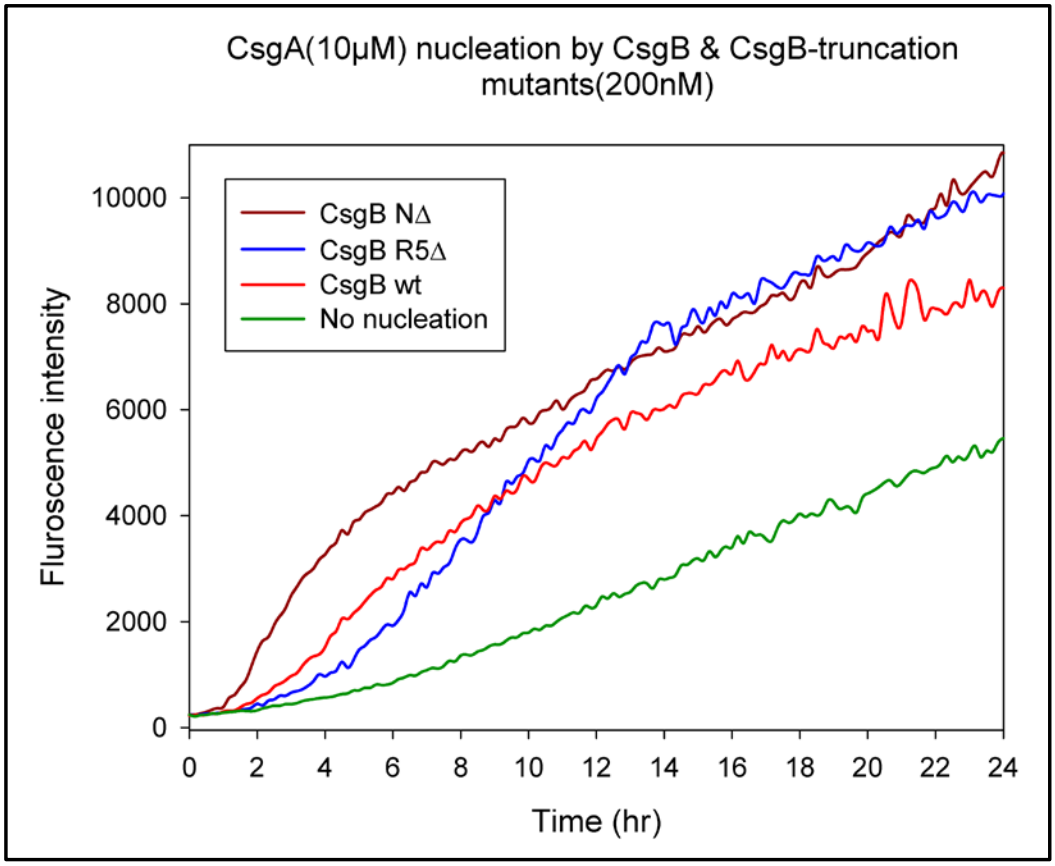


Figure 31: Nucleation of CsgA by CsgB-wt and CsgB-wt truncation mutants.

Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in presence of 200nM monomeric and soluble nucleator CsgB-wt, CsgB-ΔN & CsgB- ΔR5.Nucleation capacity in the order CsgB-ΔN > CsgB-wt > CsgB- ΔR5.

9.2.11 Construction of N and C terminal double mutant CsgB-DM (CsgB-ΔNΔR5)

To further evaluate the importance of both the N and C terminal region of the CsgB, a double mutant lacking both N and C terminal portion was cloned in the pet11d vector and purified using the same method as mentioned (material & methods).

To characterize the basic biophysical properties of the double mutant the secondary structure was studied using FTIR and CD which displayed the secondary structure rich in beta sheet (**Supplementary Figures1**) similar to wild type CsgB.

To test the amyloidogenic properties of the double mutant, self aggregation of CsgB-DM (CsgB-ΔNΔR5) was monitored using Thioflavin T binding assay **(Figure 32c**). Freshly purified CsgB-ΔNΔR5 truncation mutant in varied concentration (1, 2, 3,5,7,9 μM) was mixed with Thioflavin T simultaneously increasing the pH to 7.4, as control all the other mutants (CsgB-ΔN and CsgB-ΔR5) including CsgB-wt was used. The Thioflavin T biding showed increased fluorescence in a concentration dependent manner. The aggregation kinetics of CsgB-ΔNΔR5 displayed a lag phase of about 50-60 mins for the highest protein concentration of 9 μM but the CsgB-ΔR5 displayed a lag phase of about 150-160 mins (**Figure 32c**) while the CsgB-ΔN (**Figure 32b**) and CsgB-wt (**Figure 32a**) displayed very less or nearly no lag phase. The aggregation of the double mutant (**Figure 32c**) was faster compared to CsgB-ΔR5 mutant but slower than CsgB-ΔN and CsgB-wt. The self aggregation revealed that the double mutant displayed the hybrid aggregation property when compared to the single mutants. These results once again prove that the N terminal of the CsgB was involved in the specific nucleation while the C terminal is important for self folding of CsgB. Hence during the aggregation of double mutant the absence of N-terminal helps in initiation of the faster aggregation by self nucleation while the absence of C-terminal reduces the ability to self fold. These opposing forces allow the double mutant to display hybrid aggregation properties.

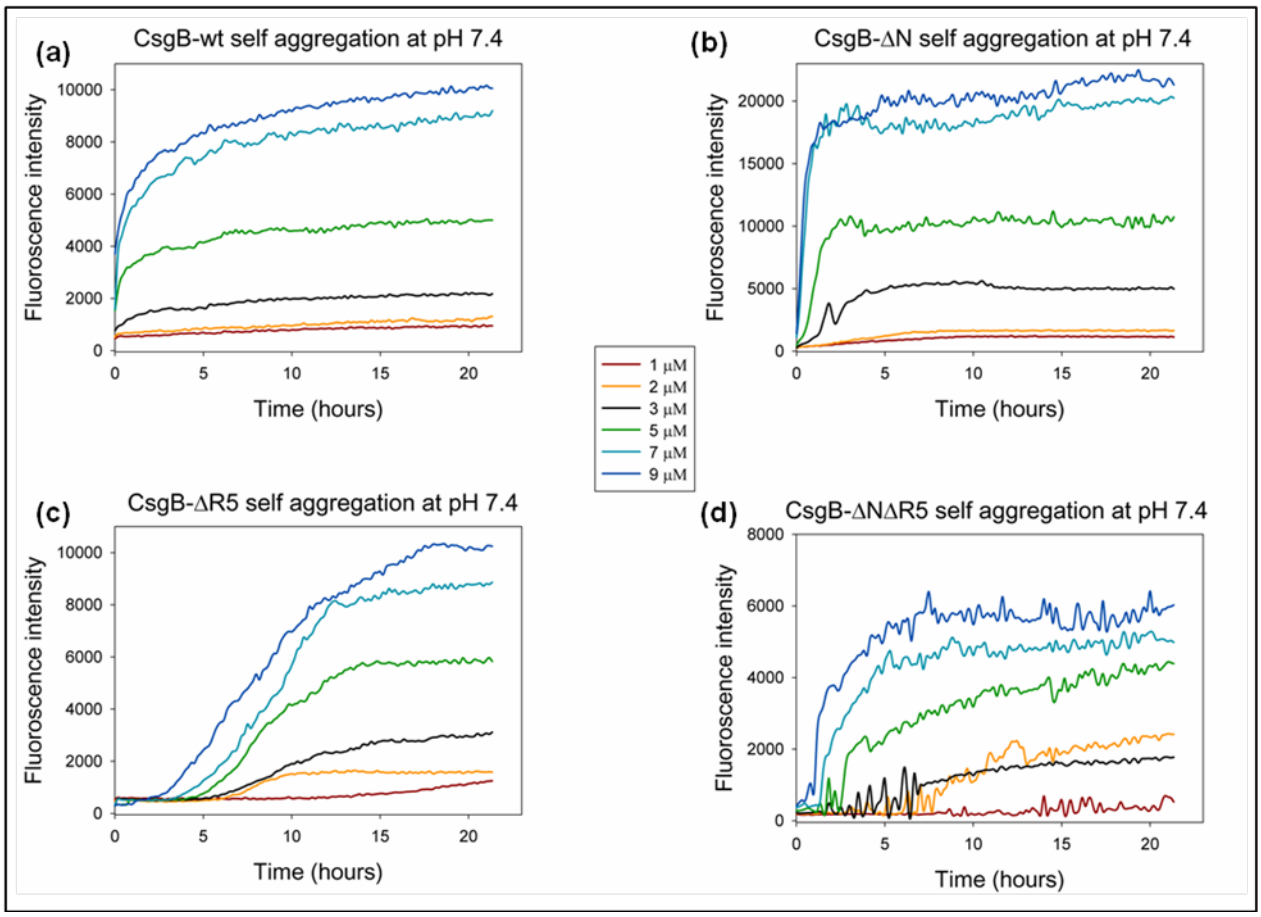


Figure 32: Aggregation kinetics monitored by Thioflavin T binding.

(a) (b) (c) (d) Fibrillation of CsgB-wt , CsgB-ΔN ,CsgB-ΔR5 & CsgB-ΔNΔR5 followed by ThT at pH 7.4. Increase in ThT fluorescence depending on time and protein concentration (1, 2, 3,5,7,9 µM). CsgB-ΔR5 shows delayed aggregation with the lag phase compared to the other CsgB-wt and the mutants.

9.2.12 CsgB-DM (CsgB-ΔNΔR5)-truncation mutant can nucleate CsgA *in vitro*

To determine the nucleator activity of CsgB-ΔNΔR5 and to evaluate the effect of varied self folding property of the double mutant, a nucleation assay of CsgA by the double mutant was set up. Freshly purified 10µM solution of CsgA was nucleated with 100nM of monomeric freshly purified CsgB-ΔNΔR5, as control all the other mutants (CsgB-ΔN and CsgB-ΔR5) including CsgB-wt was used (**Figure 33**). The lag phase of CsgA polymerization was dramatically shortened in a concentration dependent manner of the nucleator. The nucleation assay revealed that CsgB ΔN was the best nucleator followed by CsgB-wt, double mutant CsgB-ΔNΔR5 and CsgB-ΔR5 which showed the least nucleation

capability (**Figure 29,Figure 33**). These results shows that 100nM of CsgB ΔN decreased the lag phase of CsgA polymerization to 50-60 mins, CsgB-wt decreased the lag phase to 160-180 mins, CsgB ΔNΔR5 decreased the lag phase to 280-300 mins while CsgB-ΔR5 showed the least decrease of lag phase to 520-540 mins. To further confirm the above finding freshly purified 10μM solution of CsgA was nucleated with different concentrations (100nM, 200nM, 500nM) of monomeric freshly purified CsgB-ΔNΔR5, as control all the other mutants (CsgB-ΔN and CsgB-ΔR5) including CsgB-wt was used. The nucleation assay with different concentration of the CsgB mutants showed the same pattern (CsgB-ΔN > CsgB-wt > CsgB-ΔNΔR5 > CsgB-ΔR5) of decrease in the lag phase of CsgA polymerization (**Supplementary Figures 2**).This novel finding indicates that N-terminal region of CsgB is involved in specific nucleation of CsgA by giving directionality and sequence selective interface for the fiber growth, whereas C-terminal region of CsgB is involved in self-folding of CsgB from random coil to folded monomer.

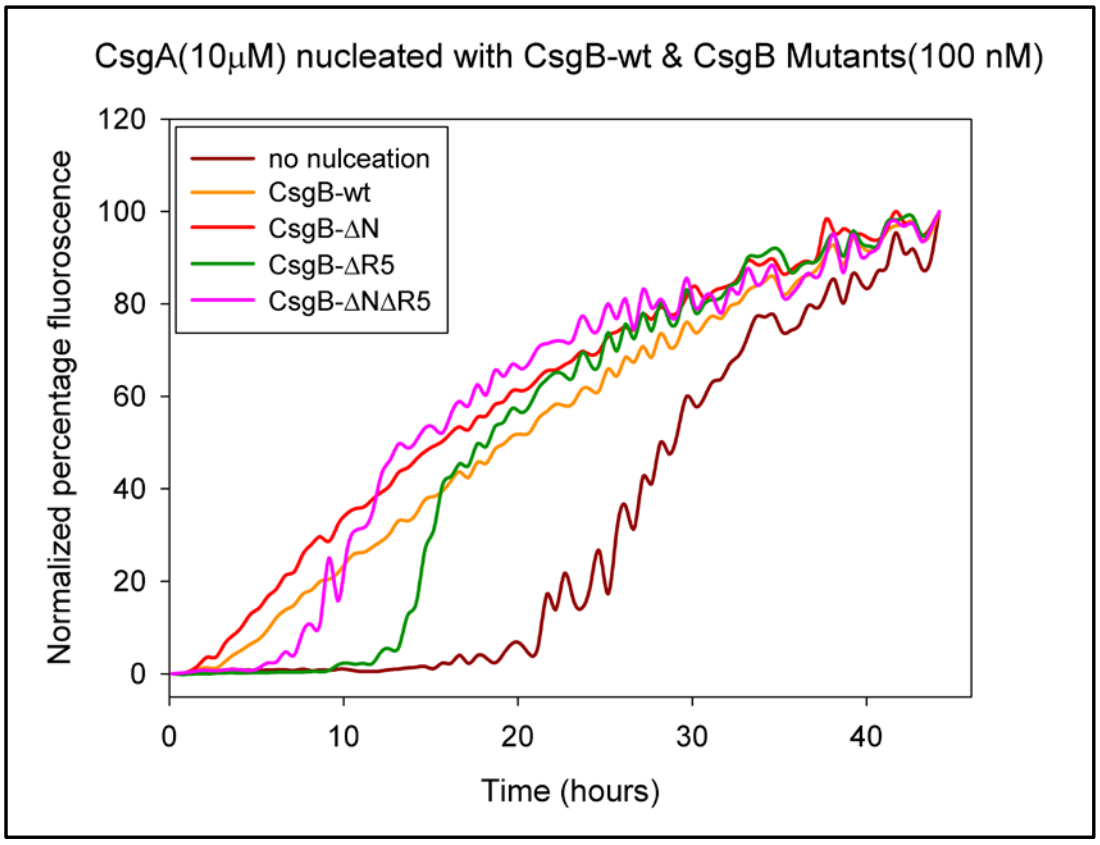


Figure 33: Nucleation of CsgA by CsgB-wt and CsgB-wt truncation mutants.

Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in presence of 100nM monomeric and soluble nucleator CsgB-wt, CsgB-ΔN, CsgB- ΔR5 & CsgB-ΔNΔR5.Nucleation capacity of the CsgB protein follows this order (CsgB-ΔN > CsgB-wt > CsgB-ΔNΔR5 > CsgB-ΔR5)

9.2.13 Sequence-Specific Analysis of Secondary Structure Elements

In order to further relate the varied nucleation capacity to the structure of CsgB-truncation mutants, sequence specific secondary structure of the truncation mutants were analyzed by HDX. Protein expression of Both CsgB-truncation mutant's ΔN and ΔR5 were carried out in a minimal media supplemented with 15NH4Cl and 13C-glucose to obtain CsgB-truncation mutants that is 15N and 13C labeled. After cultivation and cell harvest, cells were lysed under native conditions and the insoluble material was solubilised subsequently. The solubilisation was followed by IMAC under denaturing conditions using Ni-Sepharose. Eluted proteins was then spun at 50000 rpm for 2 hours at 25^0 C to remove any preformed seeds and finally desalted using Sephadex G25 in to the potassium phosphate buffer at pH 3. To ensure complete buffer exchange, conductivity was monitored. Only protein fractions devoid of denaturant were pooled and fibrillised by increasing the buffer pH to 7.4. To exclude non specific aggregates the fibrillized samples were washed with 0.1 % SDS, since the ordered fibers are resistant to the SDS wash only the non specific aggregates were dissolved by SDS. For further HDX experiments the protein samples were fibrillized at a very low concentration corresponding to approximately 30 μM and washed with SDS to enrich the ordered aggregates.

9.2.14 H/D Exchange NMR Analysis of CsgB-trunc ΔN Fibrils

Primary results with the HDX revealed higher level of heterogeneity in CsgB-ΔN truncation mutant as seen in wild type CsgB.

44 % of CsgB residues displayed biphasic behavior in the exchange plot (**Figure 34** for example amino acid 52A, 67L and 91A) indicating heterogeneity (Luhrs et al., 2005; Vilar et al., 2008). 56% of CsgB residues displayed a monoexponential decay (**Figure 34** for example amino acid 53G, 93I, & 116I) suggesting a well defined and homogeneous structure of the fibrils. The residue showing biphasic behavior indicates presence of two different types of population which were analyzed separately and plotted as mentioned (**Figure 21**). The exchange rate of the major subpopulation revealed secondary structure of CsgB-ΔN truncation mutant consisting of two β-strands that are connected by a short β-turn for R1, R2 and R4 repeating units similar to the wild type CsgB (**Figure 22**). Interestingly the secondary structure of R3 repeating unit showed only a single beta strand but the second beta strand showed heterogeneity with the major population displayed fast H/D exchange rate. Similarly, R5 repeating unit also showed major population displaying fast H/D exchange rate. The change in the secondary structure of R3 and R5 repeating units could be attributed to the truncation, but the amyloid core still shows similar secondary structure compared to the CsgB-wt. The backbone amides showing biphasic behavior constituted a subpopulation having fast H/D exchange as well as the second subpopulation displaying slow exchange rate. The exchange rate for both populations was analyzed separately. The exchange rate of the CsgB-ΔN truncation mutant was plotted based on their relative population. Exchange rate (**Figure 34**) and relative population (**Supplementary Figures 6**) of the major subpopulation was shown in green while the minor population was shown in grey.

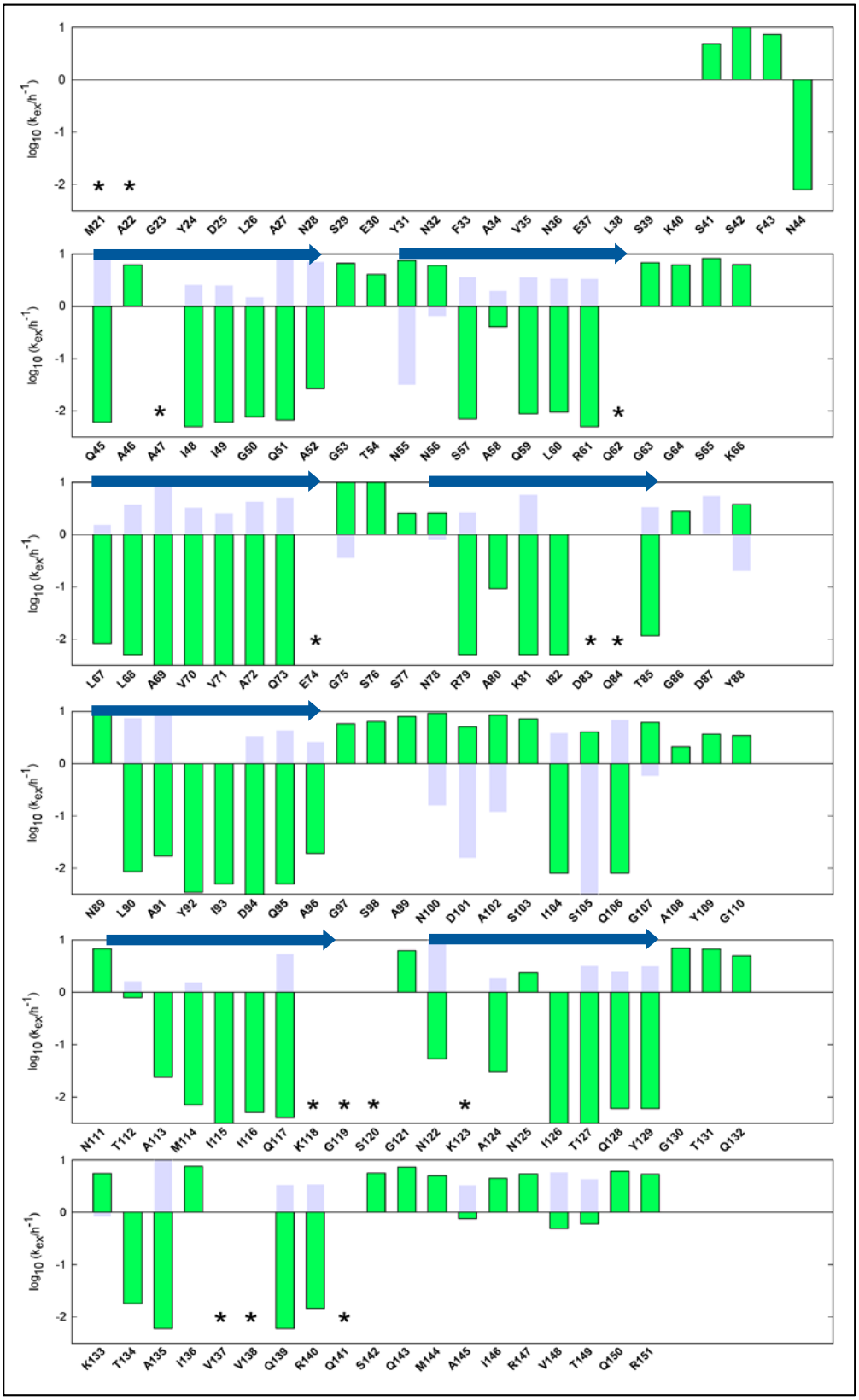


Figure 34: H/D exchange NMR data of CsgB-ΔN and secondary structure prediction.

Green and grey bars indicate exchange rates of CsgB-ΔN k_{ex} (h^{-1}). Grey bars indicate exchange rates of the minor population and green indicate major population in case of biphasic behaviour. Residues marked with a star are missing assignments. Blue arrows indicate β-strands.

9.2.15 H/D Exchange NMR Analysis of CsgB-trunc ΔR5 Fibrils

HDX revealed lower level of heterogeneity in CsgB- ΔR5 truncation mutant compared to both CsgB- ΔN and wild type CsgB indicating higher level of homogeneity in the fibers. One possible reason could be that the slower aggregation of the CsgB- ΔR5 mutant provides enough time to fold homogenously.

40 % of CsgB residues displayed biphasic behavior in the exchange plot (**Figure 35** for example amino acid 50G, 68L and 91A) indicating heterogeneity (Luhrs et al., 2005; Vilar et al., 2008). 60% of CsgB residues displayed a monoexponential decay (**Figure 35** for example amino acid 53G, 93I, & 116I) suggesting a well defined and homogeneous structure of the fibrils. The residue showing biphasic behavior was analyzed separately and plotted as mentioned (**Figure 21**). The exchange rate of the major subpopulation revealed secondary structure of CsgB-ΔR5 truncation mutant consisting of two β-strands that are connected by a short β-turn for R1, R2 and R4 repeating units similar to the wild type CsgB (**Figure 22**). Interestingly, the N-terminal region which showed additional beta strand in the secondary structure of CsgB-wt was not seen in the N-terminal region of the CsgB-ΔR5 truncation mutant but showed a region of high heterogeneity with the major population displaying fast HD exchange rate. The change in the secondary structure of R3 and R5 repeating units was similar compared to the CsgB-ΔN truncation mutant, but the amyloid core still shows similar secondary structure compared to the CsgB-wt. This change in the secondary structure of the two mutants could indicate the head to tail interaction between the different molecules of the CsgB protein in the oligomers (**Figure 36**). The backbone amides showing biphasic behavior constituted a subpopulation having fast H/D exchange as well as the second subpopulation displaying slow exchange rate. The exchange rate for both populations was analyzed separately. The exchange rate of the CsgB-ΔR5 truncation mutant was plotted against their relative population. Exchange rate (**Figure 35**) and relative population (**Supplementary Figures 7**) of the major subpopulation was shown in green while the minor population was shown in grey.

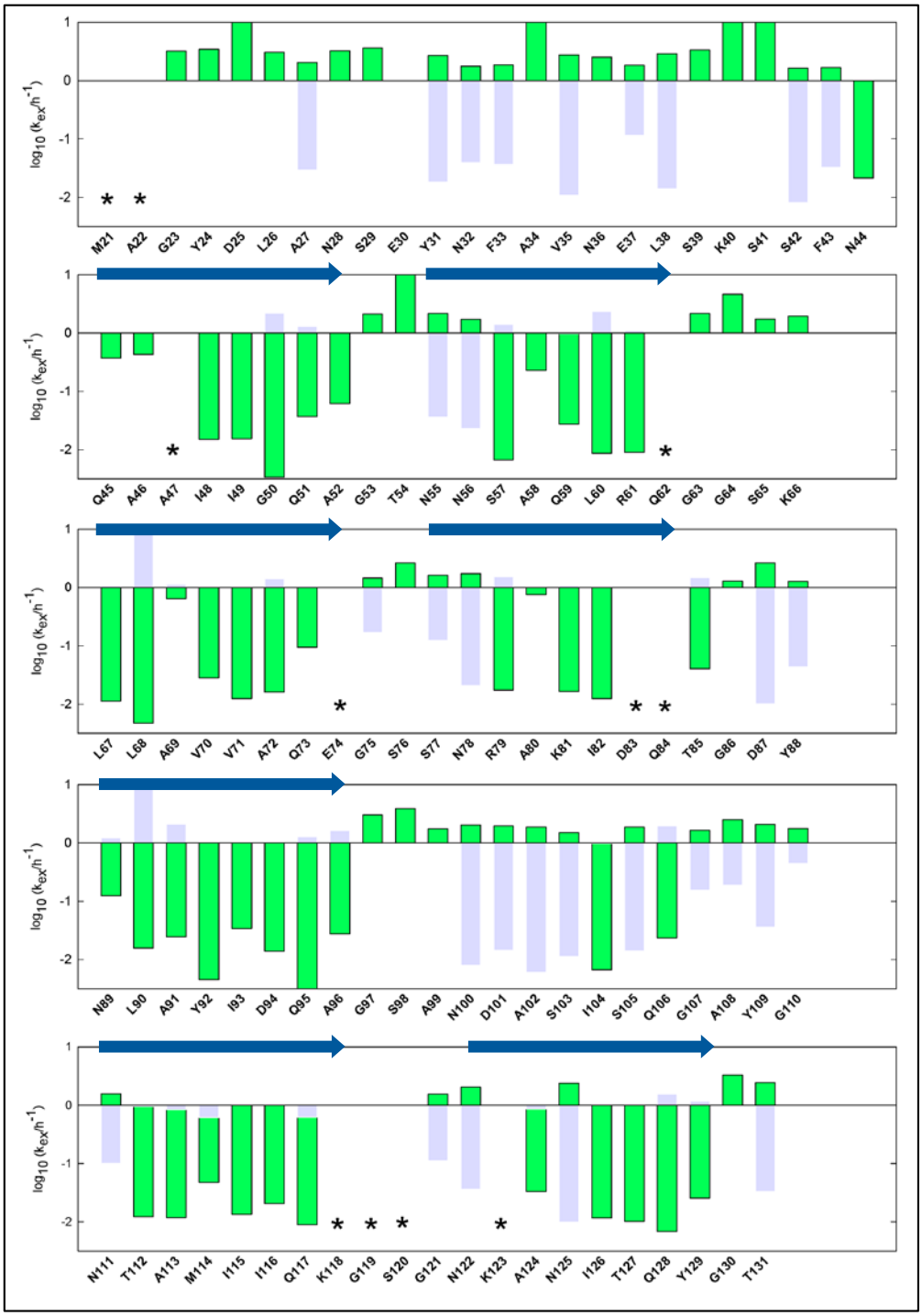


Figure 35: H/D exchange NMR data of CsgB-ΔR5 and secondary structure prediction.

Green and grey bars indicate exchange rates of CsgB-ΔN k_{ex} (h^{-1}). Grey bars indicate exchange rates of the minor population and green indicate major population in case of biphasic behaviour. Residues marked with a star are missing assignments. Blue arrows indicate β-strands.

9.2.16 Discussion

An inherent challenge in functional amyloid systems is to tightly control the process of amyloid aggregation reaction such that it does not occur at the wrong time or in the wrong place. Curli biogenesis also has to make sure that the process is robust, physiologically less demanding and economical. To achieve the above mentioned conditions curli biogenesis separates the nucleation and elongation properties into two distinct proteins (CsgB and CsgA respectively) and then dictating when and where those proteins interact. The separation of nucleation and polymerization process makes curli an efficient amyloid propagation system. In the first part of the result based on the sequence specific secondary structure of full length CsgB-wt, we hypothesized the nucleation surface would be the N-terminal which consists of an extra beta strand. This beat strand would then provide a sequence selective interface for the CsgA nucleation. The secondary structure also showed higher level of heterogeneity on the C-terminal region of repeat R5. To investigate the importance of both N terminal and C-terminal region truncation mutants were constructed based on the HDX data.

The results clearly showed the amyliodogenic properties of the truncation mutants but with varied folding properties hence influencing its nucleation property on CsgA. The aggregation of the CsgB-ΔN truncation mutant was similar to that of full length CsgB-wt with nearly no lag phase, but the CsgB-ΔR5 truncation mutant showed a delayed aggregation with the lag phase of 200-300 mins. The N-terminal truncation showed faster aggregation at pH 7.4 similar to CsgB-wt (**Figure 28**) since the pI of the N-terminal truncation mutant of CsgB was 9.90 similar to the CsgB-wt. In contrast, the C-terminal truncated CsgB showed delayed aggregation similar to CsgA-wt since the pI of the C-terminal truncation mutant of CsgB was 6.57 similar to CsgA-wt. These results indicated the variation of self aggregation in the mutants could be due to two reasons

1) Change in the iso-electric point (pI) of the protein which leads to change in the ionic interaction at pH 7 as discussed in the result-1

2) N-terminal is involved in the specific nucleation of CsgA and C-terminal was involved in the process of self folding of CsgB.

The varied folding properties was further visualized in the nucleation where CsgB-ΔN truncation mutant shown to be the most efficient nucleator followed by CsgB-wt and CsgB-ΔR5 truncation mutant. These result shows that CsgB-ΔN truncation mutant and

CsgB-wt folds faster hence nucleates the CsgA faster. In contrast, the CsgB-ΔR5 truncation mutant folds slower and hence nucleates CsgA late. The efficient nucleation capacity of the CsgB-ΔN truncation mutant could be explained as follows

1) CsgB-ΔN truncation mutant folded faster due to its higher ionic interactions influenced by its higher pI as compared to the full length CsgB-wt. However, the difference in the aggregation kinetics could not be tracked with the thioflavin T assay

2) In the absence of N-terminal extra beta strand a "specific nucleating surface", both the R1 and R5 repeat can act as nucleator providing the growing end for the CsgA attachment.

The varied nucleation of the CsgB-wt and the truncation mutants was further confirmed by monitoring over EM. The aggregation process revealed unique insight in the amyloid propagation, which showed different stages of aggregation where soluble monomer would first convert to amorphous aggregates and then these aggregates convert to short and thin proto-fibrils fibers followed by the aggregates attached to the lateral and growing end of proto-fibers, further converting to long and thick matured fibers.

The truncation mutants showed varied aggregation and nucleation property; however the secondary structure as analyzed by FTIR and also the morphology of the fibers analyzed by EM showed no difference. The self-aggregation and the nucleation results indicated that N-terminal CsgB was involved in the nucleation process and C-terminal was involved in the folding and/or aggregation of CsgB. To further verify the above mentioned properties of both the N and C terminal CsgB, a double mutant lacking both N and C terminal was constructed. CsgB-ΔNΔR5 truncation mutant showed the hybrid property of both N and C terminal truncation proteins during the aggregation kinetics and nucleation assay. This behavior could once again be explained by the ionic interactions influenced by pI of the protein which was 8.14 intermediate to that of CsgB-ΔN (pI=9.90) and CsgB-ΔR5 (pI=6.57) truncation mutants.

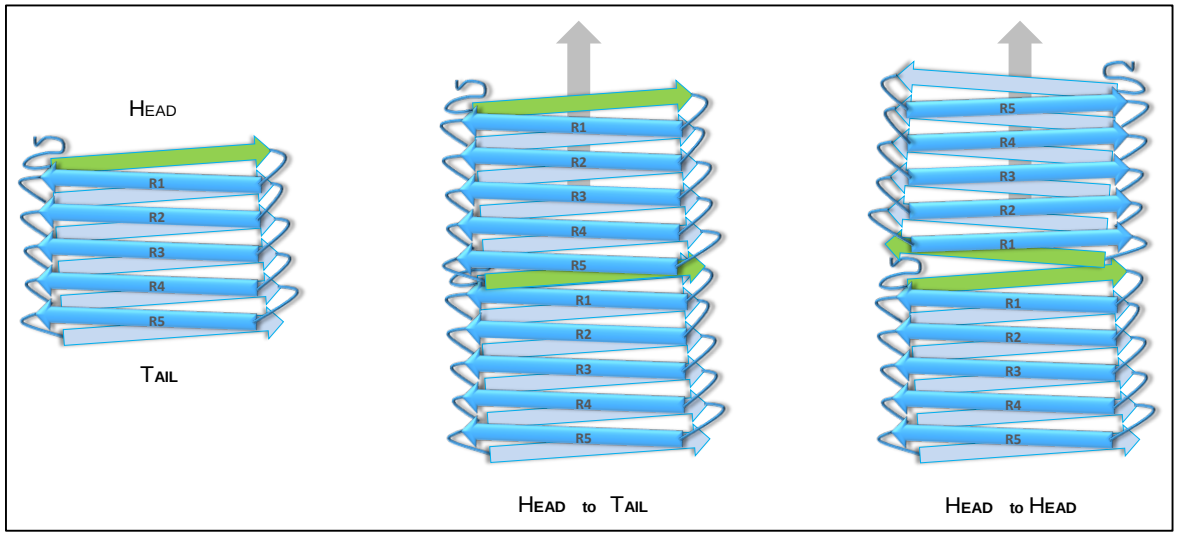


Figure 36: Possible orientation of CsgB-wt molecules in the CsgB fibers.

Folded and monomeric CsgB-wt (blue colour) shows skewed and overlapping arrangement of beta strand. The extra beta strand in the N-terminal region indicated green colour. Based on HDX data of CsgB mutants 'head to tail' orientation was possible rather than 'head to head' or 'tail to tail' orientation. Arrow indicates the growth of fiber direction.

The sequence specific secondary structure of both CsgB-ΔN and CsgB-ΔR5 truncation mutant was similar. Exchange rate of the major subpopulation revealed two β-strands that are connected by a short β-turn for each repeating units of R1, R2 and R4 similar to the CsgB-wt (**Figure 22**). Interestingly the sequence specific secondary structure of the repeating unit R3 in both of the mutants showed only a single beta strand while the second beta strand showing lot of heterogeneity with the major population displaying fast HD exchange rate (**Figure 34, Figure 35**). Interestingly the N-terminal and C-terminal region in either of the mutants showed higher heterogeneity with the major population displaying fast HD exchange rate (**Figure 34, Figure 35**). This change in the sequence specific secondary structure of both CsgB-ΔN and CsgB-ΔR5 truncation mutant compared to CsgB–wt indicates head to tail orientation (**Figure 36**) of the CsgB molecules during the fiber formation.

CsgB fibers are formed by the interaction of the N-terminal region of the one molecule of CsgB to the C-terminal region of another molecule forming oligomers; hence truncation in either N or C terminal region would lead to the loss of secondary structure in both N and C terminal region of CsgB-ΔN and CsgB-ΔR5 truncation mutants.

9.3 Chapter 3- Structural characterization of major curli protein CsgA by seeding through curli isolated from MC4100 strain

9.3.1 Abstract:

The extracellular curli proteins of Enterobacteriaceae form fibrous structures that are involved in biofilm formation and adhesion to host cells. The aim was to determine the detailed function of these structures. To this end, we examined pure culture systems using the bacterial strain MC4110 which over express natural curli. The curli was purified using the modified method published for isolation of curli (Chapman et al., 2002). The biophysical characterization of curli by FTIR indicated the presence of beta sheet rich secondary structure similar to the recombinant CsgA. In order to determine the biological function of the isolated curli, recombinant CsgA was seeded by the curli and the seeding process was monitored by Thioflavin T assays. The CsgA fibers formed when seeded with curli was then analyzed using finger print H/D exchange NMR (**Figure 41**) and compared to the CsgA fibers formed from the nucleation of CsgB and the CsgB mutants. The finger print analysis of CsgA fibers formed upon seeding with curli revealed similar H/D exchange and hence similar sequence specific secondary structure compared to the CsgA fibers formed by nucleating with CsgB and CsgB mutants. The above results indicated that CsgB provides an efficient and effective sequence selective interface for the nucleation of CsgA by fastening the aggregation process and preventing the loss or misfolding of CsgA in vivo.

9.3.2 Starting Point of Curli purification

Commonly used laboratory E. coli K-12 strain, MC4100 was used for this study (Chapman et al., 2002). MC4100 was routinely grown at 30°C on LB agar lacking salt. Media supplemented with Congo red (40 µg/ml) was used for determination of the phenotype of the bacterial strain. Based on the appearance of the colony on the Congo red agar, phenotypes of the bacterial strains were classified. **Figure 37** shows the morphotypes of the (Urinary tract infection) UTI strains grown on Congo red agar at 37° C for 24 hr. The strains expressing the rdar (red, dry and rough surface) morphotype (**Figure 37 a**) showed high biofilm-formation capacity whereas pas (pink and smooth surface) morphotype

(**Figure 37 b**) strains showed a medium capacity to form biofilms. In contrast, strains expressing the saw (smooth and white surface) morphotype displayed very low biofilm-forming capacity (**Figure 37 c**). MC4100 grown at 30°C on LB agar lacking salt, supplemented with Congo red (40 µg/ml) showed rdas (red, dry and smooth surface) morphotype.

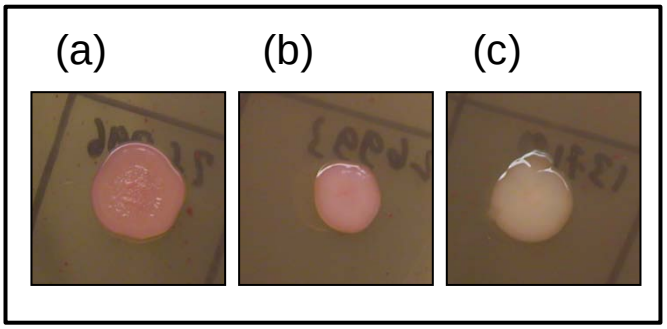


Figure 37: Morphotypes and regulation of extracellular matrix of UTI strains.

Morphotypes are: **(a)** rdar **(b)** pdas **(c)** saw. Cells were grown on LB agar with no salt supplemented with CR (40 µg/ml) for 24 h at 37 °C.

9.3.3 Isolation and purification of curli from *E. coli* K-12 strain, MC4100

For the isolation and purification of natural curli, MC4100 was spread as a lawn on 40-50 150-mm LB agar plates with no salt and grown at 30°C for 60 hours before being scraped from the plates and suspended in 300 ml of 10 mM tris (pH 7.4). Bacteria were sonicated in the water bath for 15 mins, 3 times at 10-min intervals. The bacteria were pelleted by centrifuging two times at low speed (5000*g*, 10 min). The supernatant was reconstituted with150 mM NaCl and the curli pelleted by centrifuging at 13,000*g*. The pellet was resuspended in 300 ml of 10 mM tris (pH 7.4), 150 mM NaCl, and incubated on ice for 30 min before being centrifuged at 13,000*g*. This procedure was repeated three times. The pellet was again suspended in 30 ml of 10 mM tris (pH 7.4) and centrifuged at 35,000*g*. The pellet was then suspended in 10 mM tris (pH 7.4) with 1% SDS and 2% triton X-100 at room temperature and left overnight (14 hours). The pellet was then washed with 10 mM tris (pH 7.4) to remove triton-x followed by centrifuging at 35,000g. The pellet was then mixed with an equal amount of 2× SDS loading buffer and subjected to electrophoresis for 5 hours on a 12% SDS-PAGE gel. The curli remaining in the slot after electrophoresis was

recovered followed by washing with distilled water (Chapman et al., 2002; Collinson et al., 1991).

Curli purified from MC4100 was analyzed by SDS–polyacrylamide gel electrophoresis after a brief treatment with 90% formic acid (FA) to depolymerise the Curli polymers into a ~18 -kDa protein (**Figure 38 a**).

9.3.4 Morphology of the precipitates-

The ultrastructure of Curli isolated from MC4100 was investigated by transmission electron microscopy (TEM). Freshly purified Curli revealed the presence of highly ordered, amyloid-like fibers similar to the recombinant CsgA (**Figure 38b**).

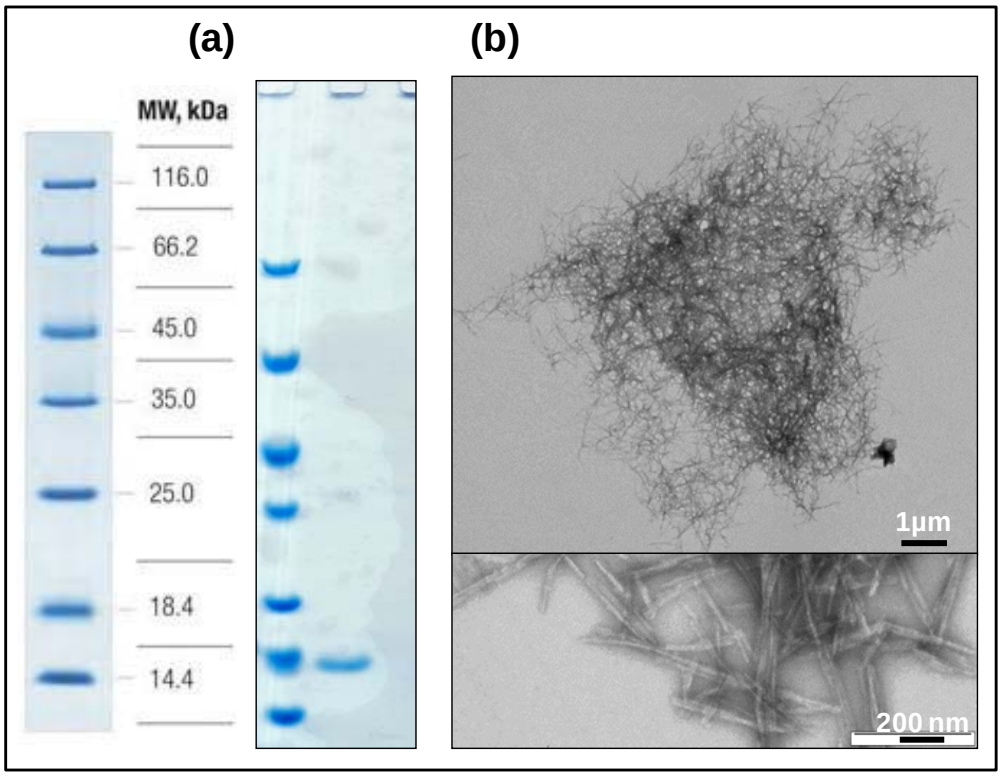


Figure 38: Biochemical property of CURLI isolated form MC4100.

(a) Coomassie stain SDS-PAGE of curli samples isolated from E. coli strain MC4100. The molecular weight of the protein is approximately 14.5 kDa, but runs at 18 kDa on the gel. Molecular size markers (in kilo daltons) are indicated on the left **(b)** Negative-stain EM micrographs of curli samples isolated from E. coli strain MC4100 grown on LB agar plates at 30°C for 60 hours. Magnification indicated on the right corner top picture (1µm) and bottom (200nm).

9.3.5 Secondary structure analysis of MC4100 isolated CURLI

The secondary structure of Curli was determined using BioATR–FTIR. The FTIR spectra showed amide I band between 1700 cm−1 and 1600 cm−1 .Amide I showed maximum absorbance at 1620 cm−1 (**Figure 39)** characteristics of amyloids fibrils (Hiramatsu and Kitagawa, 2005) . The FTIR spectrum of curli showed spectra similar to full length CsgA (**Figure 15b**) indicating the presence of similar secondary structure.

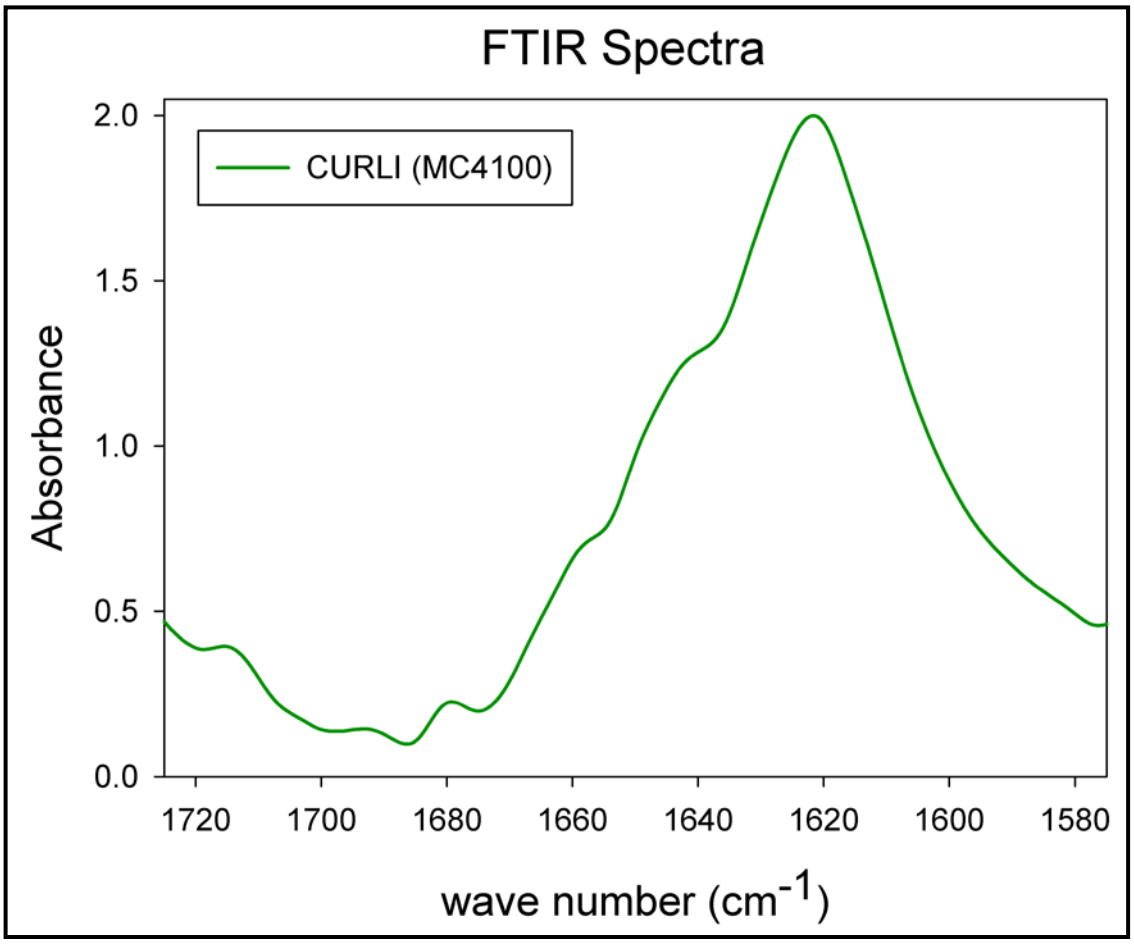


Figure 39: Biophysical characterization of CURLI isolated from MC4100.

FT-IR spectra showing the amide I band of aggregates (1 mg/ml) formed by isolated CURLI samples (green curve).

9.3.6 Curli fibers can seed recombinant CsgA in vitro

The biochemical and biophysical analysis of purified curli demonstrated its amyloidogenic properties. The next question was if the isolated curli fibers provided a seeding surface for recombinant CsgA polymerization. CsgA transition into an amyloid fiber in vitro is characterized by distinguishable lag, growth, and stationary phases. The lag phase can be

shortened by the addition of preformed CsgA fibers. The lag phase of CsgA polymerization was dramatically shortened when different concentrations (0.02, 0.5,1, and 2 %) of curli fibrillar aggregates were added to the reaction (**Figure 40**). This result indicates that the isolated curli fibers are capable of seeding soluble monomeric recombinant CsgA.

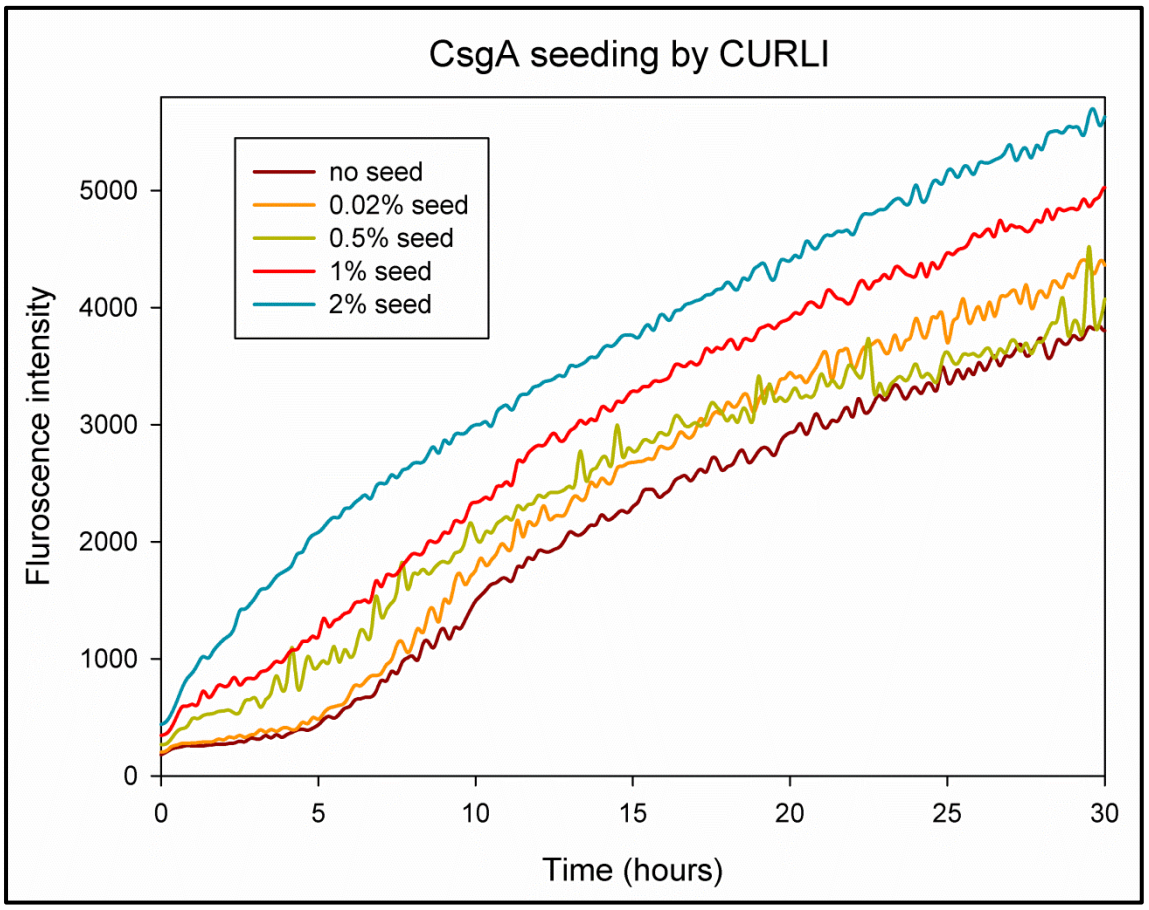


Figure 40: Curli seeding CsgA.

Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in the presence of various concentration of the isolated curli. CsgA without any seed used as control.

The isolated curli fibers displayed secondary structures rich in β-sheet content that was capable of binding with the amyloid specific dye thioflavin T and could seed soluble monomeric recombinant CsgA. To emulate the in vivo structure of curli, the purified curli was used as a template to seed CsgA (Paravastu et al., 2008; Paravastu et al., 2006; Paravastu et al., 2009; Petkova et al., 2005) because the preformed fibrils of curli can propagate their structures by seeded growth *in vitro* using recombinant CsgA. This property of self-propagating, molecular-level structural polymorphism of amyloid fibrils

underlies the phenomenon of strains in yeast prions and possibly mammalian prions (Caughey et al., 1998; Chien et al., 2004; Wickner et al., 2004). To determine the secondary structure of curli, parallel control experiments were performed on fibers formed by nucleating CsgA with full length CsgB-wt and CsgB mutants. Moreover, seeded growth from curli permits to prepare isotopically labeled fibrils in the milligram quantities required for H/D exchange NMR

9.3.7 Recombinant CsgA fibril growth by seeding with Curli

To generate seed specific recombinant CsgA fiber sample, protein expression was carried out in minimal medium supplemented with $^{15}NH_4Cl$ and 13C-glucose to obtain CsgA that was 15N and 13C labeled. After cultivation and cell harvest, cells were lysed under native conditions and the insoluble material was solubilized subsequently. The solubilisation was followed by IMAC under denaturing conditions using Ni-Sepharose. Eluted protein was then spun at 50000 rpm for 2 hours at 25° C to remove any preformed seeds and finally desalted using Sephadex G25 gel-filtration column in to the potassium phosphate buffer at pH 7.4. To ensure specific seeding CsgA was desalted at very low concentration corresponding to 30µM followed by seeding with 1% curli fibers. The seeded recombinant CsgA was fibrillized to produce "generation 1" fibrils. To exclude the unlabeled curli, further desalting of recombinant CsgA was seeded with "generation 1" fibrils.

9.3.8 Fingerprint H/D Exchange NMR Analysis of Curli seeded CsgA fibrils

15N labeled CsgA fibrils chosen for H/D exchange NMR analysis were incubated in D2O buffer for various time points (0 h, 40 min, 1.2 h, 2 h, 24 h) and studied by fingerprint H/D exchange experiments as explained in materials and methods. The H/D exchange (HDX) pattern or fingerprint for the fibers formed by CsgA seeded by Curli was similar to the H/D exchange pattern or fingerprint of fibers formed by CsgA nucleated by CsgB-wt (**Figure 41**). The peak intensity of each time point was determined by integrating the

rectangular squares covering each peak (**Figure 41** red rectangle). To differentiate the fiber polymorphism, the ratio of peak intensity between the Curli seeded CsgA and CsgB-wt nucleated CsgA fibers was calculated and plotted against the amino acid number (**Figure 42**). The ratio of peak intensity was calculated by the equation below and the value equivalent to one indicated similar H/D exchange pattern. The ratio of peak intensity for different fibers was calculated for each time point (Supplementary Figures 3, 4&5). Primary results with the fingerprint H/D exchange revealed similar sequence specific secondary structure between the Curli seeded CsgA, CsgB-wt and mutants nucleated CsgA fibers.

.

Equation1 ⟶ $CsgA_{NUCLEATED}CsgB\text{-}wt = \frac{I_B}{I_C}$

Equation2 ⟶ $CsgA_{NUCLEATED}CsgB\Delta N = \frac{I_{\Delta N}}{I_C}$

Equation3 ⟶ $CsgA_{NUCLEATED}CsgB\Delta R5 = \frac{I_{\Delta R5}}{I_C}$

Equation4 ⟶ $CsgA_{NUCLEATED}CsgB\Delta N\ \Delta R5 = \frac{I_{\Delta N\ \Delta R5}}{I_C}$

I_B = Peak intensity of CsgA fibers **nucleated** by CsgB-wt

$I_{\Delta N}$ = Peak intensity of CsgA fibers **nucleated** by CsgB-ΔN

$I_{\Delta R5}$ = Peak intensity of CsgA fibers **nucleated** by CsgB-ΔR5

$I_{\Delta N\Delta R5}$ = Peak intensity of CsgA fibers **nucleated** by CsgB-ΔNΔR5

I_C = Peak intensity of CsgA fibers **seeded** by CURLI

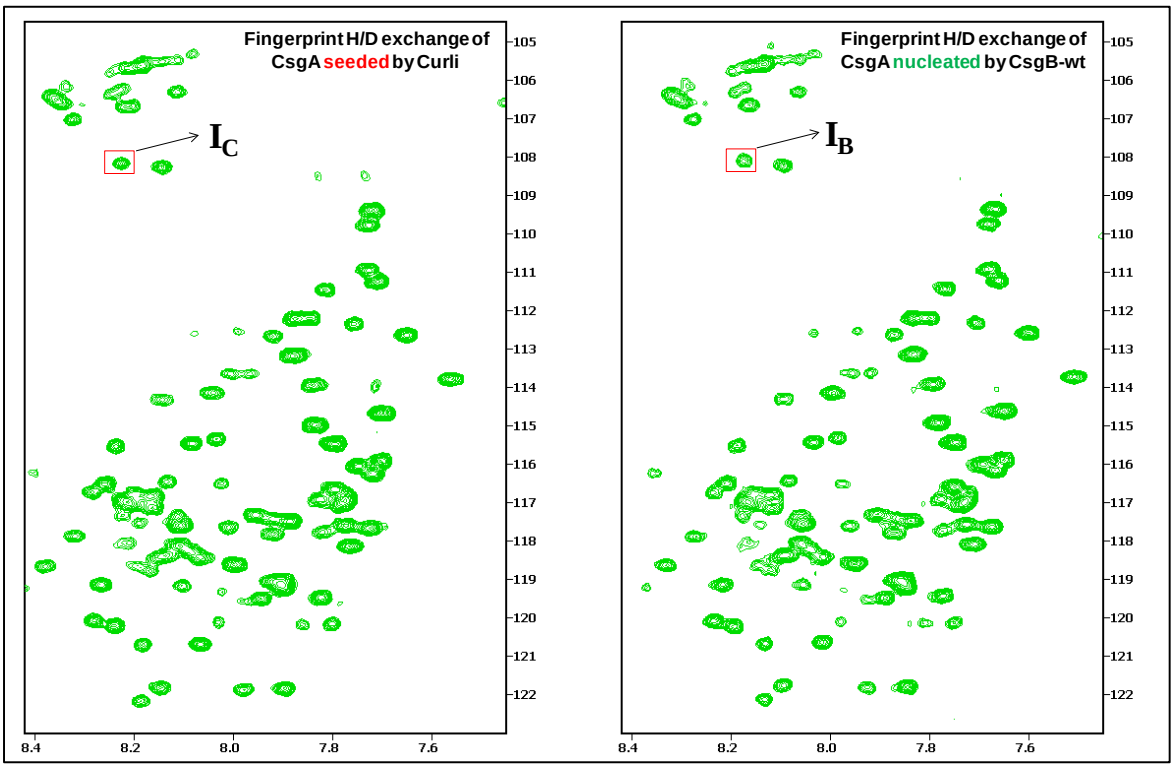


Figure 41: Finger print H/D exchange of CsgA fibers at time point 1hr20mins.

ctHSQC spectra of uniformly 15N labeled CsgA-wt (~20 mg wet pellet of fibrils) in d_6-DMSO containing 0.1 % (v/v) d_1-TFA. **Left** spectra correspond to **1hr20mins** H/D exchanged CsgA fibers formed by **seeding** with Curli and **right** spectra correspond to **1hr20mins** H/D exchanged CsgA fibers formed by **nucleating** with CsgB-wt.

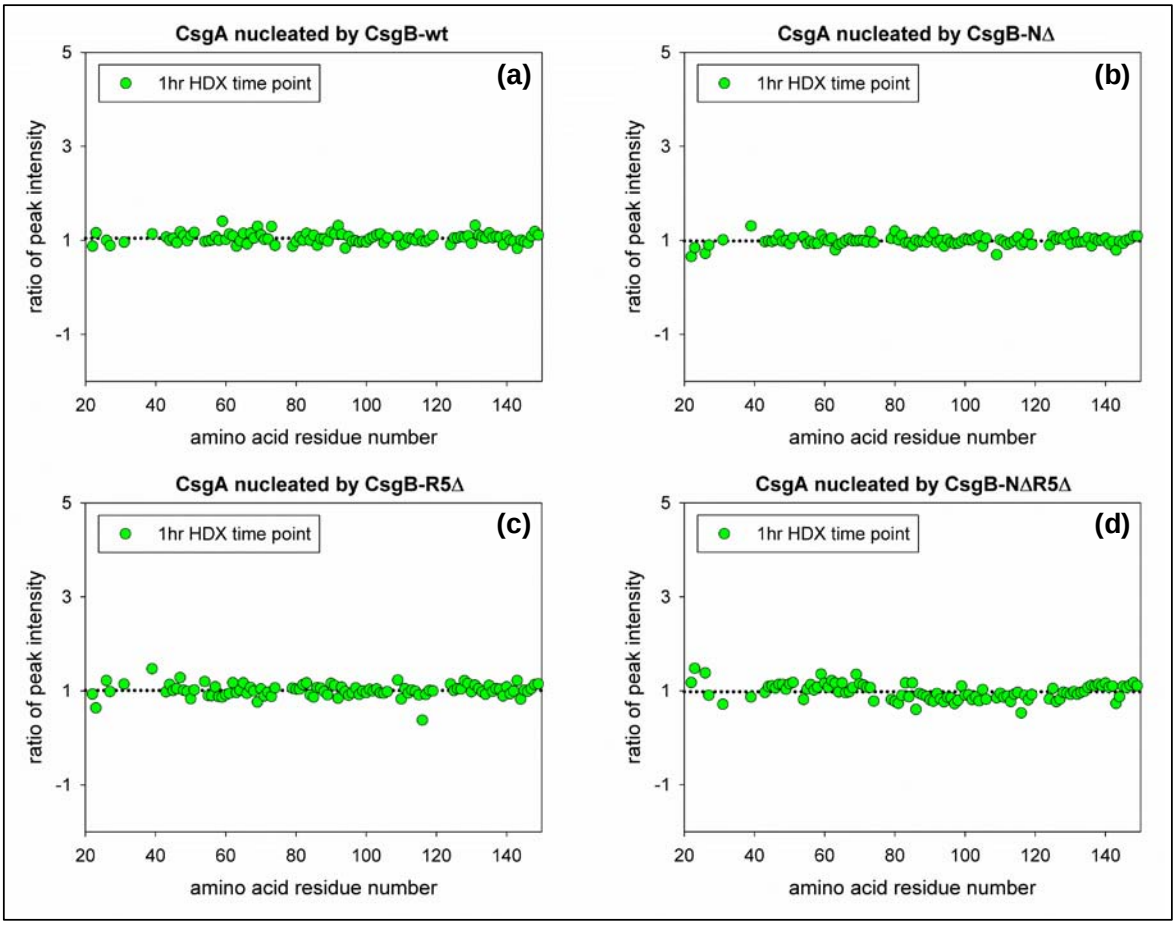


Figure 42: Ratio of peak intensity of CsgA fibers to determine polymorphism.

(a) Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-wt over fibers formed by **seeding** with Curli $\mathbf{I_B/\ I_c}$. **(b)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔN over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N}/\ I_c}$. **(c)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta R5}/\ I_c}$ **(d)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔNΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N\Delta R5}/\ I_c}$.

9.3.9 Discussion

Data presented above show that Curli extracted from MC4100 could accelerate fibril formation in recombinant CsgA in the same way that preformed recombinant CsgA fibrils could seed soluble CsgA for fibril growth (Paravastu et al., 2008; Petkova et al., 2005) **.** The sequence specific secondary structure of purely recombinant CsgA (Unpublished work Agnes Zimmer) characterized by solution NMR in previous H/D exchange studies is most likely to have similar structure as Curli seeded CsgA.

The sequence specific secondary structure of Curli seeded CsgA was similar to CsgB-wt nucleated CsgA fibers. Additionally, finger print H/D exchange revealed that CsgB-ΔN

and CsgB-ΔR5 truncation mutants nucleated CsgA fibers were similar to Curli seeded CsgA fibers.

Majority of the amino acids showed a ratio of peak intensity equal to one but few of the amino acid especially in the N-terminal region showed variation. The amino acid showing variation was located in the loop region hence showed some level of heterogeneity. Although CsgB-ΔNΔR5 truncation mutant nucleated CsgA showed increased heterogeneity in the fibers, a complete H/D exchange analysis would reveal the clear picture of fiber polymorphism.

The results clearly demonstrated that CsgB provides sequence selective interface for the CsgA nucleation by fastening the aggregation process and preventing the loss or misfolding of CsgA in vivo

10. Conclusion

Amyloid fibrils are commonly associated with diverse mammalian aliments and implicated in many deadly human diseases (**Table 1**). However an increasing number of examples of amyloid structures with no link to protein deposition diseases have been found in several lower organisms; including bacteria, fungi, and insects and subsequently in humans performing beneficial functions in nature as functional amyloid. These fibrillar aggregates have the morphological, structural, and tinctorial properties that allow them to be classified as amyloid fibrils. In contrast to disease-associated amyloids, functional amyloids are the product of regulated cellular processes which are well coordinated. Functional amyloids are indespensible for important physiological functions in the cell and provide insight into the mechanisms of protein homeostasis, folding, and misfolding (Aiba-Masago et al., 1999). Although functional amyloids were first discovered in microbes, their presence and relevance in higher organisms including humans has been identified (Aguzzi, 2008; Fowler et al., 2006; True and Lindquist, 2000) (**Table 2**). One particularly well-studied is that of the proteinaceous fibrils formed from the protein curlin that are used by Escherichia coli to colonize inert surfaces and mediate binding to host proteins. There are a growing number of examples of amyloid structures performing beneficial functions in nature as functional amyloid. The study of curli assembly machinery provides a unique template for understanding controlled amyloid propagation. Amyloidogenesis of the major curli subunit protein CsgA is dependent on the minor curli subunit protein CsgB *In vivo*, which forms as

dedicated nucleator for amyloid growth (Rochet and Lansbury, 2000a). Nucleation forms the rate-limiting step of amyloid propagation and its nature remains poorly understood.
This thesis work has been aimed at understanding the molecular details of CsgB polymerization and its function as nucleator in curli biogenesis.

10.1.1 CsgB can nucleate CsgA *in vitro*

The first chapter of the results showed the amyliodogenic properties of full length CsgB-wt including beta sheet rich secondary structure, binding to the thioflavin T an amyloid biding dye. The aggregation kinetics CsgB was faster compared to CsgA corroborating its role as a nucleator. The enhanced folding property of CsgB results in faster aggregation and vice versa, hence provides the right kind of template and avoiding the non specific random seed formation during fiber growth. This process may help in preventing formation of heterogeneous fibers and possibly toxic intermediates.
The aggregation and/or folding of CsgB is predominantly driven by ionic interaction leading to faster folding of CsgB which is highly positively charged, which would then nucleate the highly negatively charged CsgA on the cell surface. A novel finding in this part of the result showed that CsgA polymerization was shortened by soluble monomeric CsgB-wt which acts as nucleator. The above finding led to furnish a refined model of extracellular nucleation/precipitation where both CsgA and CsgB are secreted to the cell surface. CsgB folds faster compared to CsgA and nucleate CsgA to form fibers. During which single molecules of CsgB are incorporated in between the long oligomers of CsgA which was shown by immunolabelling methods (Bian and Normark, 1997). The results obtained clearly demonstrated that monomeric and soluble CsgB act as nucleator of CsgA during in vivo curli biogenesis, contrary to the previous belief that CsgB would form oligomers or fibers on the cell surface which would later provide the platform for CsgA to form fibers (Hartl et al., 2011).The secondary structure analysis of the CsgB fibers by HDX NMR showed higher level of heterogeneity. This high level of heterogeneity in the fiber structure indicates the functional structure of CsgB to be a folded monomer rather than fibers and also indicates the important role curli biogenesis proteins including CsgE, CsgF and CsgG in CsgB folding

10.1.2 N and C-terminal repeating units in CsgB govern nucleation responsiveness

An inherent challenge in functional amyloid systems is to tightly control the process of amyloid aggregation reaction such that it does not occur at the wrong time or in the wrong place. Curli biogenesis also has to make sure that the process is robust, physiological less demanding and economical. Separating the nucleation and elongation properties into two distinct proteins (CsgB and CsgA respectively) and then dictating when and where those proteins interact would help in achieving the above mentioned conditions. The separation of nucleation and polymerization process makes curli an efficient amyloid propagation system. Based on the secondary structure from the first part of the result mutation analysis was used to investigate the importance of both N terminal and C-terminal region. The results clearly showed that the truncation didn't affect the amyliodogenic properties of the mutants but could influence the folding properties hence influencing its nucleation property on CsgA. The varying folding properties of the truncation mutant were predominantly driven by ionic interaction.

The truncation analysis revealed that CsgB-ΔN truncation mutant and CsgB-wt folds faster hence nucleates the CsgA faster on contrary the CsgB-ΔR5 truncation mutant folds slower and hence nucleates CsgA late. The nucleation capacity of the CsgB-wt and the truncation mutants followed the order as - CsgB-ΔN > CsgB-wt > CsgB-ΔNΔR5 > CsgB-ΔR5.

This novel finding indicates that N-terminal region of CsgB is involved specific nucleation of CsgA by giving the directionality and sequence selective interface for the fiber growth, whereas C-terminal region of CsgB is involved in self-folding of CsgB from random coil to folded monomer.

10.1.3 Curli aggregation monitored by EM revealed different stages of aggregation process

CsgA polymerization involves a transient structurally conserved intermediate that implies a common polymerization pathway between functional and disease-associated amyloids where nucleation forms the rate-limiting step (Jarrett and Lansbury, 1993). CsgA polymerization is a triphasic process consisting of lag phase where there is formation of nucleus, log or growth phase and the stationary phase where complete conversion of

monomeric protein to fibers occur, preformed CsgA fibers or nucleator CsgB can eliminate the lag phase.

Nucleation process of CsgA with soluble monomeric CsgB and CsgB mutants was monitored through a detailed EM analysis **Figure 30**. The analysis provided a unique insight in the process of aggregation, which showed different stages of aggregation where soluble monomer would first convert to amorphous aggregates and then these aggregates convert to short and thin proto-fibrils. The aggregates further attached to the lateral and growing end of proto-fibers, further converting to long and thick matured fibers. The detailed EM analysis revealed a novel finding, displaying presence of different subpopulations of protein aggregates at different stages of the aggregation during the curli polymerization.

10.1.4 Secondary structure analysis of N and C-truncation of CsgB reveal the orientation of the single CsgB molecule in the fibers.

The sequence specific secondary structure of both CsgB-ΔN and CsgB-ΔR5 truncation mutant revealed two β-strands that are connected by a short β-turn for each repeating units of R1, R2 and R4 similar to the CsgB-wt (**Figure 22**). On contrary the sequence specific secondary structure of the repeating unit R3 in both of the mutants showed only a single beta strand while the second beta strand showing lot of heterogeneity with the major population displaying fast HD exchange rate (**Figure 34, Figure 35**). Interestingly the N-terminal and C-terminal region in either of the mutants showed higher heterogeneity with the major population displaying fast HD exchange rate (**Figure 34, Figure 35**). This change in the sequence specific secondary structure of both CsgB-ΔN and CsgB-ΔR5 truncation mutant compared to CsgB–wt indicates head to tail orientation (**Figure 36**) of the CsgB molecules during the fiber formation.

CsgB fibers are formed by the interaction of the N-terminal region of the one molecule of CsgB to the C-terminal region of another molecule forming oligomers; hence truncation in either N or C terminal region would lead to the loss of secondary structure in both N and C terminal region and also in the R3 repeat region of CsgB-ΔN and CsgB-ΔR5 truncation mutants.

The sequence specific secondary structure of both CsgB-ΔN and CsgB-ΔR5 truncation mutant was similar. Exchange rate of the major subpopulation revealed two β-strands that

are connected by a short β-turn for each repeating units of R1, R2 and R4 similar to the CsgB-wt (**Figure 22**). Interestingly the sequence specific secondary structure of the repeating unit R3 in both of the mutants showed only a single beta strand while the second beta strand showing lot of heterogeneity with the major population displaying fast HD exchange rate (**Figure 34, Figure 35**). Interestingly the N-terminal and C-terminal region in either of the mutants showed higher heterogeneity with the major population displaying fast HD exchange rate (**Figure 34, Figure 35**). This change in the sequence specific secondary structure of both CsgB-ΔN and CsgB-ΔR5 truncation mutant compared to CsgB–wt indicates head to tail orientation (**Figure 36**) of the CsgB molecules during the fiber formation.

CsgB fibers are formed by the interaction of the N-terminal region of the one molecule of CsgB to the C-terminal region of another molecule forming oligomers; hence truncation in either N or C terminal region would lead to the loss of secondary structure in both N and C terminal region of CsgB-ΔN and CsgB-ΔR5 truncation mutants.

10.1.5 *In vivo* structure of curli is similar to the CsgB nucleated CsgA fibers

The aim of the chapter 3 of the result was to determine the detailed function of the natural curli structures. To achieve the above aim pure culture systems using the bacterial strain MC4110 which over express natural curli. Curli was isolated to the purest and displayed amyliodogenic properties as analyzed by thioflavin T binding and FTIR. Isolated curli was able to seed unstructured CsgA. The fibers formed by curli seeded CsgA was analyzed by finger print HDX the sequence specific secondary structure was similar compared with the CsgA fibers formed by nucleating with CsgB and CsgB mutants. The above results indicated that CsgB provides an efficient and effective sequence selective interface for the nucleation of CsgA by fastening the aggregation process and preventing the loss or misfolding of CsgA in vivo.

11. Future Directions

In this thesis I extensively explore the nucleation process in curli. H/D exchange NMR experiments and nucleation studies revealed first residue-specific information about the minor curli subunit CsgB fibers and its role as a nucleator.

One of my most exciting discoveries was to determine the functional structure of CsgB to be a folded monomer rather than fibers. The above result raised some interesting questions on *in vivo* structure of CsgB as monomer to be determined when incorporated in the curli formed by major subunit protein CsgA.

Another novel finding was based on the nucleation analysis of the CsgB truncation mutants the importance of N and C terminal region of CsgB was determined. These results once again provide an exciting opportunity of studying the nucleation behaviour of these truncation mutants *in vivo* on the cell surface of *E.coli*.

12. References

(2006). Filling the Gaps in the A Cascade Hypothesis of Alzheimers Disease. Current Alzheimer Research *3*, 421-430.

Abrahams, C.G., and Bax, C.D. (1991). [Registration of animal drugs]. Tijdschr Diergeneeskd *116*, 371-373.

Agrawal, N.J., Kumar, S., Wang, X., Helk, B., Singh, S.K., and Trout, B.L. (2011). Aggregation in protein-based biotherapeutics: computational studies and tools to identify aggregation-prone regions. J Pharm Sci *100*, 5081-5095.

Aguzzi, A. (2008). Unraveling prion strains with cell biology and organic chemistry. Proceedings of the National Academy of Sciences of the United States of America *105*, 11-12.

Ahmad, Y., Gertz, R.E., Jr., Li, Z., Sakota, V., Broyles, L.N., Van Beneden, C., Facklam, R., Shewmaker, P.L., Reingold, A., Farley, M.M., *et al.* (2009). Genetic relationships deduced from emm and multilocus sequence typing of invasive Streptococcus dysgalactiae subsp. equisimilis and S. canis recovered from isolates collected in the United States. J Clin Microbiol *47*, 2046-2054.

Aiba-Masago, S., Baba, S., Li, R.Y., Shinmura, Y., Kosugi, I., Arai, Y., Nishimura, M., and Tsutsui, Y. (1999). Murine cytomegalovirus immediate-early promoter directs astrocyte-specific expression in transgenic mice. Am J Pathol *154*, 735-743.

Arnold, U., and Ulbrich-Hofmann, R. (1999). Quantitative protein precipitation from guanidine hydrochloride-containing solutions by sodium deoxycholate/trichloroacetic acid. Analytical biochemistry *271*, 197-199.

Arnqvist, A., Olsen, A., Pfeifer, J., Russell, D.G., and Normark, S. (1992). The Crl protein activates cryptic genes for curli formation and fibronectin binding in Escherichia coli HB101. Molecular microbiology *6*, 2443-2452.

Assadi-Porter, F.M., Abildgaard, F., Blad, H., and Markley, J.L. (2003). Correlation of the Sweetness of Variants of the Protein Brazzein with Patterns of Hydrogen Bonds Detected by NMR Spectroscopy. Journal of Biological Chemistry *278*, 31331-31339.

Astbury, W.T., Dickinson, S., and Bailey, K. (1935). The X-ray interpretation of denaturation and the structure of the seed globulins. Biochem J *29*, 2351-2360 2351.

Ausubel, F.M., Brent, R., Kingston, R.E., Moor, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (2007). Current protocols in molecular biology. John Wiley & Sons Inc.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins *17*, 75-86.

Barak, J.D., Gorski, L., Naraghi-Arani, P., and Charkowski, A.O. (2005). Salmonella enterica virulence genes are required for bacterial attachment to plant tissue. Applied and environmental microbiology *71*, 5685-5691.

Barlow, D.E., Dickinson, G.H., Orihuela, B., Kulp, J.L., 3rd, Rittschof, D., and Wahl, K.J. (2010). Characterization of the adhesive plaque of the barnacle Balanus amphitrite: amyloid-like nanofibrils are a major component. Langmuir : the ACS journal of surfaces and colloids *26*, 6549-6556.

Barnhart, M.M., and Chapman, M.R. (2006a). Curli biogenesis and function. Annual review of microbiology *60*, 131-147.

Barnhart, M.M., and Chapman, M.R. (2006b). Curli biogenesis and function. Annual review of microbiology *60*, 131-147.

Barnhart, M.M., Lynem, J., and Chapman, M.R. (2006). GlcNAc-6P levels modulate the expression of Curli fibers by Escherichia coli. Journal of bacteriology *188*, 5212-5219.

Bax, A., Griffey, R.H., and Hawkins, B.L. (1983). Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR* 1. Journal of Magnetic Resonance (1969) *55*, 301-315.

Bax, A., and Ikura, M. (1991). An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. Journal of biomolecular NMR *1*, 99-104.

Bax, N.M., and Hofstee, A.W. (1991). [Swallowed foreign articles in children]. Ned Tijdschr Geneeskd *135*, 209-211.

Bensadoun, A., and Weinstein, D. (1976). Assay of proteins in the presence of interfering materials. Anal Biochem *70*, 241-250.

Berg JM, T.J., Stryer L (2002). Biochemistry.

Berthomieu, C., and Hienerwadel, R. (2009). Fourier transform infrared (FTIR) spectroscopy. Photosynthesis research *101*, 157-170.

Bian, Z., Brauner, A., Li, Y., and Normark, S. (2000). Expression of and cytokine activation by Escherichia coli curli fibers in human sepsis. The Journal of infectious diseases *181*, 602-612.

Bian, Z., and Normark, S. (1997). Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli. The EMBO journal *16*, 5827-5836.

Bian, Z., Yan, Z.Q., Hansson, G.K., Thoren, P., and Normark, S. (2001). Activation of inducible nitric oxide synthase/nitric oxide by curli fibers leads to a fall in blood pressure during systemic Escherichia coli infection in mice. The Journal of infectious diseases *183*, 612-619.

Bonar, L., Cohen, A.S., and Skinner, M.M. (1969). Characterization of the amyloid fibril as a cross-beta protein. Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine *131*, 1373-1375.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature *416*, 507-511.

Carulla, N., Caddy, G.L., Hall, D.R., Zurdo, J., Gairí, M., Feliz, M., Giralt, E., Robinson, C.V., and Dobson, C.M. (2005). Molecular recycling within amyloid fibrils. Nature *436*, 554-558.

Caughey, B., Raymond, G.J., and Bessen, R.A. (1998). Strain-dependent differences in beta-sheet conformations of abnormal prion protein. The Journal of biological chemistry *273*, 32230-32235.

Chan, H.S., and Dill, K.A. (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Proteins *30*, 2-33.

Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., and Hultgren, S.J. (2002). Role of Escherichia coli curli operons in directing amyloid fiber formation. Science *295*, 851-855.

Chen, M., Margittai, M., Chen, J., and Langen, R. (2007). Investigation of alpha-synuclein fibril structure by site-directed spin labeling. The Journal of biological chemistry *282*, 24970-24979.

Chiba, T., Hagihara, Y., Higurashi, T., Hasegawa, K., Naiki, H., and Goto, Y. (2003). Amyloid fibril formation in the context of full-length protein: effects of proline mutations on the amyloid fibril formation of beta2-microglobulin. The Journal of biological chemistry *278*, 47016-47024.

Chien, P., Weissman, J.S., and DePace, A.H. (2004). Emerging principles of conformation-based prion inheritance. Annual review of biochemistry *73*, 617-656.

Chiti, F., and Dobson, C.M. (2006a). Protein misfolding, functional amyloid, and human disease. Annual review of biochemistry *75*, 333-366.

Chiti, F., and Dobson, C.M. (2006b). Protein misfolding, functional amyloid, and human disease. Annual review of biochemistry *75*, 333-366.

Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C.M. (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proceedings of the National Academy of Sciences of the United States of America *96*, 3590-3594.

Coligan, J.E. (2003). Short Protocols in protein science. A compendium of methods from Current protocols in protein science. John Wiley & Sons Inc.

Collinson, S.K., Clouthier, S.C., Doran, J.L., Banser, P.A., and Kay, W.W. (1996). Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae. Journal of bacteriology *178*, 662-667.

Collinson, S.K., Doig, P.C., Doran, J.L., Clouthier, S., Trust, T.J., and Kay, W.W. (1993). Thin, aggregative fimbriae mediate binding of Salmonella enteritidis to fibronectin. Journal of bacteriology *175*, 12-18.

Collinson, S.K., Emody, L., Muller, K.H., Trust, T.J., and Kay, W.W. (1991). Purification and characterization of thin, aggregative fimbriae from Salmonella enteritidis. Journal of bacteriology *173*, 4773-4781.

Collinson, S.K., Parker, J.M., Hodges, R.S., and Kay, W.W. (1999a). Structural predictions of AgfA, the insoluble fimbrial subunit of *Salmonella* thin aggregative fimbriae. Journal of molecular biology *290*, 741-756.

Collinson, S.K., Parker, J.M., Hodges, R.S., and Kay, W.W. (1999b). Structural predictions of AgfA, the insoluble fimbrial subunit of Salmonella thin aggregative fimbriae. Journal of molecular biology *290*, 741-756.

Debnath, S., Chatterjee, S., Arif, M., Kundu, T.K., and Roy, S. (2011). Peptide-protein interactions suggest that acetylation of lysines 381 and 382 of p53 is important foa positive coactivator 4/p53 interaction. Journal of Biological Chemistry.

Der-Sarkissian, A., Jao, C.C., Chen, J., and Langen, R. (2003). Structural organization of alpha-synuclein fibrils studied by site-directed spin labeling. The Journal of biological chemistry *278*, 37530-37535.

Diercks, T., Coles, M., and Kessler, H. (1999). An efficient strategy for cross-peaks in 3D heteronuclear NOESY experiments. Journal of Biomolecular NMR *15*, 177-180.

Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. Nature structural biology *4*, 10-19.

Drechsler, N., Fröbel, J., Jahreis, G., Gopalswamy, M., Balbach, J., Bosse-Doenecke, E., and Rudolph, R. (2011). Binding specificity of the ectodomain of the parathyroid hormone receptor. Biophysical Chemistry *154*, 66-72.

Dueholm, M.S., Nielsen, S.B., Hein, K.L., Nissen, P., Chapman, M., Christiansen, G., Nielsen, P.H., and Otzen, D.E. (2011). Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation. Biochemistry *50*, 8281-8290.

Dueholm, M.S., Petersen, S.V., Sonderkaer, M., Larsen, P., Christiansen, G., Hein, K.L., Enghild, J.J., Nielsen, J.L., Nielsen, K.L., Nielsen, P.H., *et al.* (2010). Functional amyloid in Pseudomonas. Molecular microbiology.

Edman, P., and Begg, G. (1967). A protein sequenator. Eur J Biochem *1*, 80-91.

Englander, S., Mayne, L., Bai, Y., and Sosnick, T. (1997). Hydrogen exchange: The modern legacy of Linderstrøm Lang. Protein Science *6*, 1101-1109.

Englander, S.W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. Annu Rev Biophys Biomol Struct *29*, 213-238.

Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., and Kelly, J.W. (2006). Functional amyloid formation within mammalian tissue. PLoS biology *4*, e6.

Fowler, D.M., Koulov, A.V., Balch, W.E., and Kelly, J.W. (2007a). Functional amyloid--from bacteria to humans. Trends in biochemical sciences *32*, 217-224.

Fowler, D.M., Koulov, A.V., Balch, W.E., and Kelly, J.W. (2007b). Functional amyloid - from bacteria to humans. Trends in biochemical sciences *32*, 217-224.

Gardiennet, C., Henry, B., Kuad, P., Spiess, B., and Tekely, P. (2005). Straightforward detection of the secondary ionisation of the phosphate group and pK determinations by high-resolution solid-state 31P NMR. Chemical communications, 180-182.

Geddes, A.J., Parker, K.D., Atkins, E.D., and Beighton, E. (1968). "Cross-beta" conformation in proteins. Journal of molecular biology *32*, 343-358.

Gerstel, U., and Romling, U. (2001). Oxygen tension and nutrient starvation are major signals that regulate agfD promoter activity and expression of the multicellular morphotype in Salmonella typhimurium. Environmental microbiology *3*, 638-648.

Gerstel, U., and Romling, U. (2003). The csgD promoter, a control unit for biofilm formation in Salmonella typhimurium. Research in microbiology *154*, 659-667.

Ghiso, J., and Frangione, B. (2002). Amyloidosis and Alzheimer's disease. Advanced drug delivery reviews *54*, 1539-1551.

Gibson, D.L., White, A.P., Rajotte, C.M., and Kay, W.W. (2007). AgfC and AgfE facilitate extracellular thin aggregative fimbriae synthesis in Salmonella Enteritidis. Microbiology *153*, 1131-1140.

Glabe, C.G., and Kayed, R. (2006). Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. Neurology *66*, S74-78.

Glenner, G.G., Cuatrecasas, P., Isersky, C., Bladen, H.A., and Eanes, E.D. (1969). Physical and chemical properties of amyloid fibers. II. Isolation of a unique protein constituting the major component from human splenic amyloid fibril concentrates. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society *17*, 769-780.

Glenner, G.G., Ein, D., Eanes, E.D., Bladen, H.A., Terry, W., and Page, D.L. (1971a). Creation of "amyloid" fibrils from Bence Jones proteins in vitro. Science *174*, 712-714.

Glenner, G.G., Page, D., Isersky, C., Harada, M., Cuatrecasas, P., Eanes, E.D., DeLellis, R.A., Bladen, H.A., and Keiser, H.R. (1971b). Murine amyloid fibril protein: isolation, purification and characterization. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society *19*, 16-28.

Gophna, U., Barlev, M., Seijffers, R., Oelschlager, T.A., Hacker, J., and Ron, E.Z. (2001). Curli fibers mediate internalization of Escherichia coli by eukaryotic cells. Infection and immunity *69*, 2659-2665.

Greenwald, J., and Riek, R. (2010). Biology of amyloid: structure, function, and regulation. Structure *18*, 1244-1260.

Grzesiek, S., and Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. Journal of magnetic resonance *96*, 432-440.

Guntert, P., Dotsch, V., Wider, C., and Wüthrich, K. (1992). Processing of multi-dimensional NMR data with new software PROSA. Journal of Biomolecular NMR *2*, 619-629.

Hammar, M., Arnqvist, A., Bian, Z., Olsen, A., and Normark, S. (1995a). Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12. Molecular microbiology *18*, 661-670.

Hammar, M., Arnqvist, A., Bian, Z., Olsen, A., and Normark, S. (1995b). Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12. Molecular microbiology *18*, 661-670.

Hammar, M., Bian, Z., and Normark, S. (1996). Nucleator-dependent intercellular assembly of adhesive curli organelles in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *93*, 6562-6566.

Hammer, N.D., Schmidt, J.C., and Chapman, M.R. (2007a). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. Proceedings of the National Academy of Sciences of the United States of America *104*, 12494-12499.

Hammer, N.D., Schmidt, J.C., and Chapman, M.R. (2007b). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. Proceedings of the National Academy of Sciences of the United States of America *104*, 12494-12499.

Hammer, N.D., Wang, X., McGuffie, B.A., and Chapman, M.R. (2008). Amyloids: friend or foe? Journal of Alzheimer's disease : JAD *13*, 407-419.

Harper, J.D., and Lansbury, P.T., Jr. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. Annual review of biochemistry *66*, 385-407.

Harper, J.D., Lieber, C.M., and Lansbury, P.T., Jr. (1997). Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-beta protein. Chem Biol *4*, 951-959.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324-332.

Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. Nature Structural & Molecular Biology *16*, 574-581.

Hartley, D.M., Walsh, D.M., Ye, C.P., Diehl, T., Vasquez, S., Vassilev, P.M., Teplow, D.B., and Selkoe, D.J. (1999). Protofibrillar intermediates of amyloid beta-protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience *19*, 8876-8884.

Hawe, A., Sutter, M., and Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. Pharmaceutical research *25*, 1487-1499.

Heim, M., Keerl, D., and Scheibel, T. (2009). Spider silk: from soluble protein to extraordinary fiber. Angew Chem Int Ed Engl *48*, 3584-3596.

Henderson, R. (2004). Realizing the potential of electron cryo-microscopy. Quarterly reviews of biophysics *37*, 3-13.

Herczenik, E., and Gebbink, M.F.B.G. (2008). Molecular and cellular aspects of protein misfolding and disease. The FASEB Journal *22*, 2115.

Hiramatsu, H., and Kitagawa, T. (2005). FT-IR approaches on amyloid fibril structure. Biochimica et biophysica acta *1753*, 100-107.

Hofmann, H., Weininger, U., Lo¨w, C., Golbik, R.P., Balbach, J., and Ulbrich-Hofmann, R. (2008). Fast Amide Proton Exchange Reveals Close Relation between Native-State Dynamics and Unfolding Kinetics. Journal of the American Chemical Society *131*, 140-146.

Hornak, J.P. (1997-2011). The Basics of NMR.

Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H., and Goto, Y. (2002). Mapping the core of the β2-microglobulin amyloid fibril by H/D exchange. Nature structural biology *9*, 332-336.

Howie, A.J., and Brewer, D.B. (2009). Optical properties of amyloid stained by Congo red: history and mechanisms. Micron *40*, 285-301.

Hvidt, A., and Linderstrøm-Lang, K. (1954). Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. Biochimica et biophysica acta *14*, 574.

Ippel, J.H., Olofsson, A., Schleucher, J., Lundgren, E., and Wijmenga, S.S. (2002). Probing solvent accessibility of amyloid fibrils by solution NMR spectroscopy. Proceedings of the National Academy of Sciences of the United States of America *99*, 8648-8653.

Ivankov, D.N., and Finkelstein, A.V. (2004). Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. Proceedings of the National Academy of Sciences of the United States of America *101*, 8942-8944.

Jahn, T.R., Makin, O.S., Morris, K.L., Marshall, K.E., Tian, P., Sikorski, P., and Serpell, L.C. (2010). The common architecture of cross-beta amyloid. Journal of molecular biology *395*, 717-727.

Jarrett, J.T., and Lansbury, P.T., Jr. (1993). Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? Cell *73*, 1055-1058.

Jeter, C., and Matthysse, A.G. (2005). Characterization of the binding of diarrheagenic strains of E. coli to plant surfaces and the role of curli in the interaction of the bacteria with alfalfa sprouts. Molecular plant-microbe interactions : MPMI *18*, 1235-1242.

Jimenez, J.L., Guijarro, J.I., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M., and Saibil, H.R. (1999). Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. The EMBO journal *18*, 815-821.

Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. (2002). The protofilament structure of insulin amyloid fibrils. Proceedings of the National Academy of Sciences of the United States of America *99*, 9196-9201.

Jordal, P.B., Dueholm, M.S., Larsen, P., Petersen, S.V., Enghild, J.J., Christiansen, G., Hojrup, P., Nielsen, P.H., and Otzen, D.E. (2009). Widespread abundance of functional

bacterial amyloid in mycolata and other gram-positive bacteria. Applied and environmental microbiology *75*, 4101-4110.

Kayed, R., Head, E., Thompson, J.L., McIntire, T.M., Milton, S.C., Cotman, C.W., and Glabe, C.G. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science *300*, 486-489.

Keller, R. (2004). The Computer Aided Resonance Assignment Tutorial. CANTINA Verlag, Goldau, Switzerland.

Kelly, J.W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Current opinion in structural biology *8*, 101-106.

Kheterpal, I., Cook, K.D., and Wetzel, R. (2006). Hydrogen/deuterium exchange mass spectrometry analysis of protein aggregates. Methods in enzymology *413*, 140-166.

Kheterpal, I., Williams, A., Murphy, C., Bledsoe, B., and Wetzel, R. (2001). Structural features of the Aβ amyloid fibril elucidated by limited proteolysis. Biochemistry *40*, 11757-11767.

Kikuchi, T., Mizunoe, Y., Takade, A., Naito, S., and Yoshida, S. (2005). Curli fibers are required for development of biofilm architecture in Escherichia coli K-12 and enhance bacterial adherence to human uroepithelial cells. Microbiology and immunology *49*, 875-884.

Kim, I.C., Franzblau, C., Shirahama, T., and Cohen, A.S. (1969). The effect of papain, pronase, Nagarse and trypsin on isolated amyloid fibrils. Biochim Biophys Acta *181*, 465-467.

Kim, Y.S., Randolph, T.W., Manning, M.C., Stevens, F.J., and Carpenter, J.F. (2003). Congo red populates partially unfolded states of an amyloidogenic protein to enhance aggregation and amyloid fibril formation. The Journal of biological chemistry *278*, 10842-10850.

Knowles, T.P., Smith, J.F., Craig, A., Dobson, C.M., and Welland, M.E. (2006). Spatial persistence of angular correlations in amyloid fibrils. Physical review letters *96*, 238301.

Kramer, E.B., and Farabaugh, P.J. (2007). The frequency of translational misreading errors in E. coli is largely determined by tRNA competition. Rna *13*, 87-96.

Kurland, C.G. (1992). Translational accuracy and the fitness of bacteria. Annu Rev Genet *26*, 29-50.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lambert, M.P., Barlow, A.K., Chromy, B.A., Edwards, C., Freed, R., Liosatos, M., Morgan, T.E., Rozovsky, I., Trommer, B., Viola, K.L., *et al.* (1998). Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. Proceedings of the National Academy of Sciences of the United States of America *95*, 6448-6453.

Larsen, P., Nielsen, J.L., Otzen, D., and Nielsen, P.H. (2008). Amyloid-like adhesins produced by floc-forming and filamentous bacteria in activated sludge. Applied and environmental microbiology *74*, 1517-1526.

Larson, J., Lynch, G., Games, D., and Seubert, P. (1999). Alterations in synaptic transmission and long-term potentiation in hippocampal slices from young and aged PDAPP mice. Brain research *840*, 23-35.

LeVine, H., 3rd (1993). Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. Protein science : a publication of the Protein Society *2*, 404-410.

LeVine, H., 3rd (1999a). Quantification of beta-sheet amyloid fibril structures with thioflavin T. Methods in enzymology *309*, 274-284.

LeVine, H., 3rd (1999b). Quantification of beta-sheet amyloid fibril structures with thioflavin T. Methods in enzymology *309*, 274-284.

Li, R., and Woodward, C. (1999). The hydrogen exchange core and protein folding. Protein Sci *8*, 1571-1590.

Linser, R., Dasari, M., Hiller, M., Higman, V., Fink, U., Lopez del Amo, J.-M., Markovic, S., Handel, L., Kessler, B., Schmieder, P., *et al.* (2011). Proton-Detected Solid-State NMR Spectroscopy of Fibrillar and Membrane Proteins. Angewandte Chemie International Edition *50*, 4508-4512.

Liwang, A.C., and Bax, A. (1997). Solution NMR Characterization of Hydrogen Bonds in a Protein by Indirect Measurement of Deuterium Quadrupole Couplings. Journal of magnetic resonance *127*, 54-64.

Loferer, H., Hammar, M., and Normark, S. (1997). Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG. Molecular microbiology *26*, 11-23.

Lorenzo, A., and Yankner, B.A. (1994). Beta-amyloid neurotoxicity requires fibril formation and is inhibited by congo red. Proceedings of the National Academy of Sciences of the United States of America *91*, 12243-12247.

Lue, L.F., Kuo, Y.M., Roher, A.E., Brachova, L., Shen, Y., Sue, L., Beach, T., Kurth, J.H., Rydel, R.E., and Rogers, J. (1999). Soluble amyloid beta peptide concentration as a predictor of synaptic change in Alzheimer's disease. Am J Pathol *155*, 853-862.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D., and Riek, R. (2005). 3D structure of Alzheimer's amyloid-β(1-42) fibrils. Proc Natl Acad Sci U S A *102*, 17342-17347.

Maji, S.K., Perrin, M.H., Sawaya, M.R., Jessberger, S., Vadodaria, K., Rissman, R.A., Singru, P.S., Nilsson, K.P., Simon, R., Schubert, D., *et al.* (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. Science *325*, 328-332.

Malisauskas, M., Ostman, J., Darinskas, A., Zamotin, V., Liutkevicius, E., Lundgren, E., and Morozova-Roche, L.A. (2005). Does the cytotoxic effect of transient amyloid

oligomers from common equine lysozyme in vitro imply innate amyloid toxicity? The Journal of biological chemistry *280*, 6269-6275.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., and Bax, A. (1989). Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. Journal of the American Chemical Society *111*, 1515-1517.

Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. Journal of biomolecular NMR *20*, 71-75.

McLean, C.A., Cherny, R.A., Fraser, F.W., Fuller, S.J., Smith, M.J., Beyreuther, K., Bush, A.I., and Masters, C.L. (1999). Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. Annals of neurology *46*, 860-866.

Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N., and Fandrich, M. (2009). Aβ(1-40) fibril polymorphism implies diverse interaction patterns in amyloid fibrils. Journal of molecular biology *386*, 869-877.

Moechars, D., Dewachter, I., Lorent, K., Reverse, D., Baekelandt, V., Naidu, A., Tesseur, I., Spittaels, K., Haute, C.V., Checler, F., *et al.* (1999). Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. The Journal of biological chemistry *274*, 6483-6492.

Mori, S., Abeygunawardana, C., Johnson, M., and Vanzijl, P. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. Journal of Magnetic Resonance, Series B *108*, 94-98.

Naiki, H., and Gejyo, F. (1999). Kinetic analysis of amyloid fibril formation. Methods in enzymology *309*, 305-318.

Naiki, H., Hasegawa, K., Yamaguchi, I., Nakamura, H., Gejyo, F., and Nakakuki, K. (1998). Apolipoprotein E and antioxidants have different mechanisms of inhibiting Alzheimer's beta-amyloid fibril formation in vitro. Biochemistry *37*, 17882-17889.

Naiki, H., Higuchi, K., Hosokawa, M., and Takeda, T. (1989). Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. Analytical biochemistry *177*, 244-249.

Nazabal, A., and Schmitter, J.M. (2006). Hydrogen-deuterium exchange analyzed by matrix-assisted laser desorption-ionization mass spectrometry and the HET-s prion model. Methods in enzymology *413*, 167-181.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-beta spine of amyloid-like fibrils. Nature *435*, 773-778.

O'Nuallain, B., Williams, A.D., Westermark, P., and Wetzel, R. (2004). Seeding specificity in amyloid growth induced by heterologous fibrils. The Journal of biological chemistry *279*, 17490-17499.

Olsen, A., Jonsson, A., and Normark, S. (1989). Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. Nature *338*, 652-655.

Otzen, D. (2010). Functional amyloid: turning swords into plowshares. Prion *4*, 256-264.

Otzen, D., and Nielsen, P.H. (2008). We find them here, we find them there: functional bacterial amyloid. Cellular and molecular life sciences : CMLS *65*, 910-927.

Paravastu, A.K., Leapman, R.D., Yau, W.M., and Tycko, R. (2008). Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. Proceedings of the National Academy of Sciences of the United States of America *105*, 18349-18354.

Paravastu, A.K., Petkova, A.T., and Tycko, R. (2006). Polymorphic fibril formation by residues 10-40 of the Alzheimer's beta-amyloid peptide. Biophysical journal *90*, 4618-4629.

Paravastu, A.K., Qahwash, I., Leapman, R.D., Meredith, S.C., and Tycko, R. (2009). Seeded growth of beta-amyloid fibrils from Alzheimer's brain-derived fibrils produces a distinct fibril structure. Proceedings of the National Academy of Sciences of the United States of America *106*, 7443-7448.

Pauling, L., and Corey, R.B. (1951). Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. Proceedings of the National Academy of Sciences of the United States of America *37*, 729-740.

Petkova, A.T., Leapman, R.D., Guo, Z., Yau, W.M., Mattson, M.P., and Tycko, R. (2005). Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils. Science *307*, 262-265.

Pike, C.J., Walencewicz, A.J., Glabe, C.G., and Cotman, C.W. (1991). In vitro aging of beta-amyloid protein causes peptide aggregation and neurotoxicity. Brain research *563*, 311-314.

Ramirez-Alvarado, M., Merkel, J.S., and Regan, L. (2000). A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. Proceedings of the National Academy of Sciences of the United States of America *97*, 8979-8984.

Ranjbar, B., and Gill, P. (2009). Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses A Review. Chemical Biology & Drug Design *74*, 101-120.

Ritter, C., Maddelein, M.L., Siemer, A.B., Luhrs, T., Ernst, M., Meier, B.H., Saupe, S.J., and Riek, R. (2005). Correlation of structural elements and infectivity of the HET-s prion. Nature *435*, 844-848.

Robinson, L.S., Ashman, E.M., Hultgren, S.J., and Chapman, M.R. (2006). Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. Molecular microbiology *59*, 870-881.

Rochet, J.-C., and Lansbury, P.T. (2000a). Amyloid fibrillogenesis: themes and variations. Current opinion in structural biology *10*, 60-68.

Rochet, J.C., and Lansbury, P.T., Jr. (2000b). Amyloid fibrillogenesis: themes and variations. Current opinion in structural biology *10*, 60-68.

Romling, U., Sierralta, W.D., Eriksson, K., and Normark, S. (1998). Multicellular and aggregative behaviour of Salmonella typhimurium strains is controlled by mutations in the agfD promoter. Molecular microbiology *28*, 249-264.

Ryu, J.H., Kim, H., and Beuchat, L.R. (2004). Attachment and biofilm formation by Escherichia coli O157:H7 on stainless steel as influenced by exopolysaccharide production, nutrient availability, and temperature. Journal of food protection *67*, 2123-2131.

Sachse, C., Fandrich, M., and Grigorieff, N. (2008). Paired beta-sheet structure of an Aβ(1-40) amyloid fibril revealed by electron microscopy. Proceedings of the National Academy of Sciences of the United States of America *105*, 7462-7466.

Sambrook, J., and Russell, D.W. (2000). Molecular cloning - a laboratory manual. Cold Spring Harbor Laboratory Press.

Sanchez, L., Madurga, S., Pukala, T., Vilaseca, M., Lopez-Iglesias, C., Robinson, C.V., Giralt, E., and Carulla, N. (2011). Abeta40 and Abeta42 amyloid fibrils exhibit distinct molecular recycling properties. J Am Chem Soc *133*, 6505-6508.

Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T., *et al.* (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature *447*, 453-457.

Scheibel, T., Parthasarathy, R., Sawicki, G., Lin, X.M., Jaeger, H., and Lindquist, S.L. (2003). Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. Proceedings of the National Academy of Sciences of the United States of America *100*, 4527-4532.

Schmidt, M., Sachse, C., Richter, W., Xu, C., Fandrich, M., and Grigorieff, N. (2009). Comparison of Alzheimer Aβ(1-40) and Aβ(1-42) amyloid fibrils reveals similar protofilament structures. Proceedings of the National Academy of Sciences of the United States of America *106*, 19813-19818.

Schnabel, J. (2010). Protein folding: The dark side of proteins. Nature *464*, 828-829.

Schulenburg, C., Lo¨w, C., Weininger, U., Mrestani-Klaus, C., Hofmann, H., Balbach, J., Ulbrich-Hofmann, R., and Arnold, U. (2009). The Folding Pathway of Onconase Is Directed by a Conserved Intermediate. Biochemistry *48*, 8449-8457.

Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., Moslehi, J.J., Serpell, L., Arnsdorf, M.F., and Lindquist, S.L. (2000). Nucleated conformational conversion and the replication of conformational information by a prion determinant. Science *289*, 1317-1321.

Shammas, C., Donarski, J.A., and Ramesh, V. (2007). NMR structure of the peptidyl transferase RNA inhibitor antibiotic amicetin. Magnetic Resonance in Chemistry *45*, 133-141.

Shewmaker, F., McGlinchey, R.P., Thurber, K.R., McPhie, P., Dyda, F., Tycko, R., and Wickner, R.B. (2009). The functional curli amyloid is not based on in-register parallel beta-sheet structure. The Journal of biological chemistry *284*, 25065-25076.

Shirahama, T., and Cohen, A.S. (1967). Reconstitution of amyloid fibrils from alkaline extracts. J Cell Biol *35*, 459-464.

Sipe, J.D., and Cohen, A.S. (2000). Review: history of the amyloid fibril. Journal of structural biology *130*, 88-98.

Sirangelo, I., Malmo, C., Iannuzzi, C., Mezzogiorno, A., Bianco, M.R., Papa, M., and Irace, G. (2004). Fibrillogenesis and cytotoxic activity of the amyloid-forming apomyoglobin mutant W7FW14F. The Journal of biological chemistry *279*, 13183-13189.

Sousa, M.M., Cardoso, I., Fernandes, R., Guimaraes, A., and Saraiva, M.J. (2001). Deposition of transthyretin in early stages of familial amyloidotic polyneuropathy: evidence for toxicity of nonfibrillar aggregates. Am J Pathol *159*, 1993-2000.

Stryer, L. (1966). Biochemistry, 4 edn (New York: W. H. Freeman and Company).

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., and Blake, C.C. (1997a). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. Journal of molecular biology *273*, 729-739.

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., and Blake, C.C.F. (1997b). Common core structure of amyloid fibrils by synchrotron X-ray diffraction1. Journal of molecular biology *273*, 729-739.

Torok, M., Milton, S., Kayed, R., Wu, P., McIntire, T., Glabe, C.G., and Langen, R. (2002). Structural and dynamic features of Alzheimer's Aβ peptide in amyloid fibrils studied by site-directed spin labeling. The Journal of biological chemistry *277*, 40810-40815.

Townsend, M., Shankar, G.M., Mehta, T., Walsh, D.M., and Selkoe, D.J. (2006). Effects of secreted oligomers of amyloid beta-protein on hippocampal synaptic plasticity: a potent role for trimers. The Journal of physiology *572*, 477-492.

Toyama, B.H., Kelly, M.J.S., Gross, J.D., and Weissman, J.S. (2007). The structural basis of yeast prion strain variants. Nature *449*, 233-237.

Toyama, B.H., and Weissman, J.S. (2011a). Amyloid Structure: Conformational Diversity and Consequences. Annual review of biochemistry *80*, 557-585.

Toyama, B.H., and Weissman, J.S. (2011b). Amyloid structure: conformational diversity and consequences. Annual review of biochemistry *80*, 557-585.

True, H.L., and Lindquist, S.L. (2000). A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature *407*, 477-483.

Tycko, R. (2011). Solid State NMR Studies of Amyloid Fibril Structure. Annual review of physical chemistry *62*.

Uhlich, G.A., Cooke, P.H., and Solomon, E.B. (2006). Analyses of the red-dry-rough phenotype of an Escherichia coli O157:H7 strain and its role in biofilm formation and resistance to antibacterial agents. Applied and environmental microbiology *72*, 2564-2572.

Uhlich, G.A., Keen, J.E., and Elder, R.O. (2001). Mutations in the csgD promoter associated with variations in curli expression in certain strains of Escherichia coli O157:H7. Applied and environmental microbiology *67*, 2367-2370.

Valentine, R.C., Shapiro, B.M., and Stadtman, E.R. (1968). Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from Escherichia coli. Biochemistry *7*, 2143-2152.

Vassar, P.S., and Culling, C.F. (1959). Fluorescent stains, with special reference to amyloid and connective tissues. Archives of pathology *68*, 487-498.

Vidal, O., Longin, R., Prigent-Combaret, C., Dorel, C., Hooreman, M., and Lejeune, P. (1998). Isolation of an Escherichia coli K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new ompR allele that increases curli expression. Journal of bacteriology *180*, 2442-2449.

Vilar, M., Chou, H.T., Luhrs, T., Maji, S.K., Riek-Loher, D., Verel, R., Manning, G., Stahlberg, H., and Riek, R. (2008). The fold of alpha-synuclein fibrils. Proceedings of the National Academy of Sciences of the United States of America *105*, 8637-8642.

Wang, J., Dickson, D.W., Trojanowski, J.Q., and Lee, V.M. (1999). The levels of soluble versus insoluble brain Abeta distinguish Alzheimer's disease from normal and pathologic aging. Experimental neurology *158*, 328-337.

Wang, X., Hammer, N.D., and Chapman, M.R. (2008). The Molecular Basis of Functional Bacterial Amyloid Polymerization and Nucleation. Journal of Biological Chemistry *283*, 21530-21539.

Wang, X., Rochon, M., Lamprokostopoulou, A., Lunsdorf, H., Nimtz, M., and Romling, U. (2006). Impact of biofilm matrix components on interaction of commensal Escherichia coli with the gastrointestinal cell line HT-29. Cellular and molecular life sciences : CMLS *63*, 2352-2363.

Wasmer, C., Soragni, A., Sabate, R., Lange, A., Riek, R., and Meier, B.H. (2008). Infectious and noninfectious amyloids of the HET-s(218-289) prion have different NMR spectra. Angew Chem Int Ed Engl *47*, 5839-5841.

West, M.W., Wang, W., Patterson, J., Mancias, J.D., Beasley, J.R., and Hecht, M.H. (1999). De novo amyloid proteins from designed combinatorial libraries. Proceedings of the National Academy of Sciences of the United States of America *96*, 11211-11216.

White, A.P., Collinson, S.K., Banser, P.A., Gibson, D.L., Paetzel, M., Strynadka, N.C., and Kay, W.W. (2001). Structure and characterization of AgfB from Salmonella enteritidis thin aggregative fimbriae. Journal of molecular biology *311*, 735-749.

Wickner, R.B., Edskes, H.K., Roberts, B.T., Baxa, U., Pierce, M.M., Ross, E.D., and Brachmann, A. (2004). Prions: proteins as genes and infectious entities. Genes & development *18*, 470-485.

Williams, A.D., Portelius, E., Kheterpal, I., Guo, J.T., Cook, K.D., Xu, Y., and Wetzel, R. (2004). Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis. Journal of molecular biology *335*, 833-842.

Wiltzius, J.J., Landau, M., Nelson, R., Sawaya, M.R., Apostol, M.I., Goldschmidt, L., Soriaga, A.B., Cascio, D., Rajashankar, K., and Eisenberg, D. (2009). Molecular mechanisms for protein-encoded inheritance. Nature Structural & Molecular Biology *16*, 973-978.

Wiltzius, J.J., Sievers, S.A., Sawaya, M.R., Cascio, D., Popov, D., Riekel, C., and Eisenberg, D. (2008). Atomic structure of the cross-beta spine of islet amyloid polypeptide (amylin). Protein science : a publication of the Protein Society *17*, 1467-1474.

Wittekind, M., and Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-and beta-carbon resonances in proteins. Journal of magnetic resonance Series B *101*, 201-205.

Wolfe, L.S., Calabrese, M.F., Nath, A., Blaho, D.V., Miranker, A.D., and Xiong, Y. (2010). Protein-induced photophysical changes to the amyloid indicator dye thioflavin T. Proceedings of the National Academy of Sciences of the United States of America *107*, 16863-16868.

Wood, S.J., Wetzel, R., Martin, J.D., and Hurle, M.R. (1995). Prolines and amyloidogenicity in fragments of the Alzheimer's peptide Aβ. Biochemistry *34*, 724-730.

Wüthrich, K. (1990). Protein structure determination in solution by NMR spectroscopy. Journal of Biological Chemistry *265*, 22059-22062.

Xia, Y., Kong, Y., Thomsen, T.R., and Halkjaer Nielsen, P. (2008). Identification and ecophysiological characterization of epiphytic protein-hydrolyzing saprospiraceae ("Candidatus Epiflobacter" spp.) in activated sludge. Applied and environmental microbiology *74*, 2229-2238.

Zandomeneghi, G., Krebs, M.R., McCammon, M.G., and Fandrich, M. (2004). FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. Protein science : a publication of the Protein Society *13*, 3314-3321.

Zhang, R., Hu, X., Khant, H., Ludtke, S.J., Chiu, W., Schmid, M.F., Frieden, C., and Lee, J.M. (2009). Interprotofilament interactions between Alzheimer's Aβ1-42 peptides in amyloid fibrils revealed by cryoEM. Proceedings of the National Academy of Sciences of the United States of America *106*, 4653-4658.

Appendix

A) Amino acid sequences

CsgA (curli subunit) [*E. coli*] – amino acid sequence (GenBank entry AAA23616)

```
  1 MKLLKVAAIA AIVFSGSALA GVVPQYGGGG NHGGGGNNSG PNSELNIYQY GGGNSALALQ
 61 TDARNSDLTI TQHGGGNGAD VGQGSDDSSI DLTQRGFGNS ATLDQWNGKN SEMTVKQFGG
121 GNGAAVDQTA SNSSVNVTQV GFGNNATAHQ Y
```

CsgB (nucleation component of curli monomers) [*E. coli*]– amino acid sequence (GenBank entry CAA62281)

```
  1 MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL
 61 RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQGAYG NTAMIIQKGS
121 GNKANITQYG TQKTAIVVQR QSQMAIRVTQ R
```

Recombinant expression constructs

All constructs carried a hexa histidine tag, highlighted in blue.

$CsgA_{21\text{-}151}$ – amino acid sequence

```
 20 MGVVPQYGGG GNHGGGGNNS GPNSELNIYQ YGGGNSALAL QTDARNSDLT ITQHGGGNGA
 80 DVGQGSDDSS IDLTQRGFGN SATLDQWNGK NSEMTVKQFG GGNGAAVDQT ASNSSVNVTQ
140 VGFGNNATAH QYHHHHHH
```

$CsgB_{21\text{-}151}$ – amino acid sequence

```
 20 MAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL RQGGSKLLAV VAQEGSSNRA
 80 KIDQTGDYNL AYIDQAGSAN DASISQGAYG NTAMIIQKGS GNKANITQYG TQKTAIVVQR
140 QSQMAIRVTQ RHHHHHH
```

CsgB-ΔN – amino acid sequence

```
 40 SSFNQAAIIG QAGTNNSAQL RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN
 80 DASISQGAYG NTAMIIQKGS GNKANITQYG TQKTAIVVQR QSQMAIRVTQ RHHHHHH
```

CsgB-ΔR5 – amino acid sequence

```
 20 MAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL RQGGSKLLAV VAQEGSSNRA
 80 KIDQTGDYNL AYIDQAGSAN DASISQGAYG NTAMIIQKGS GNKANITQYG THHHHHH
```

CsgB- ΔNΔR5 – amino acid sequence

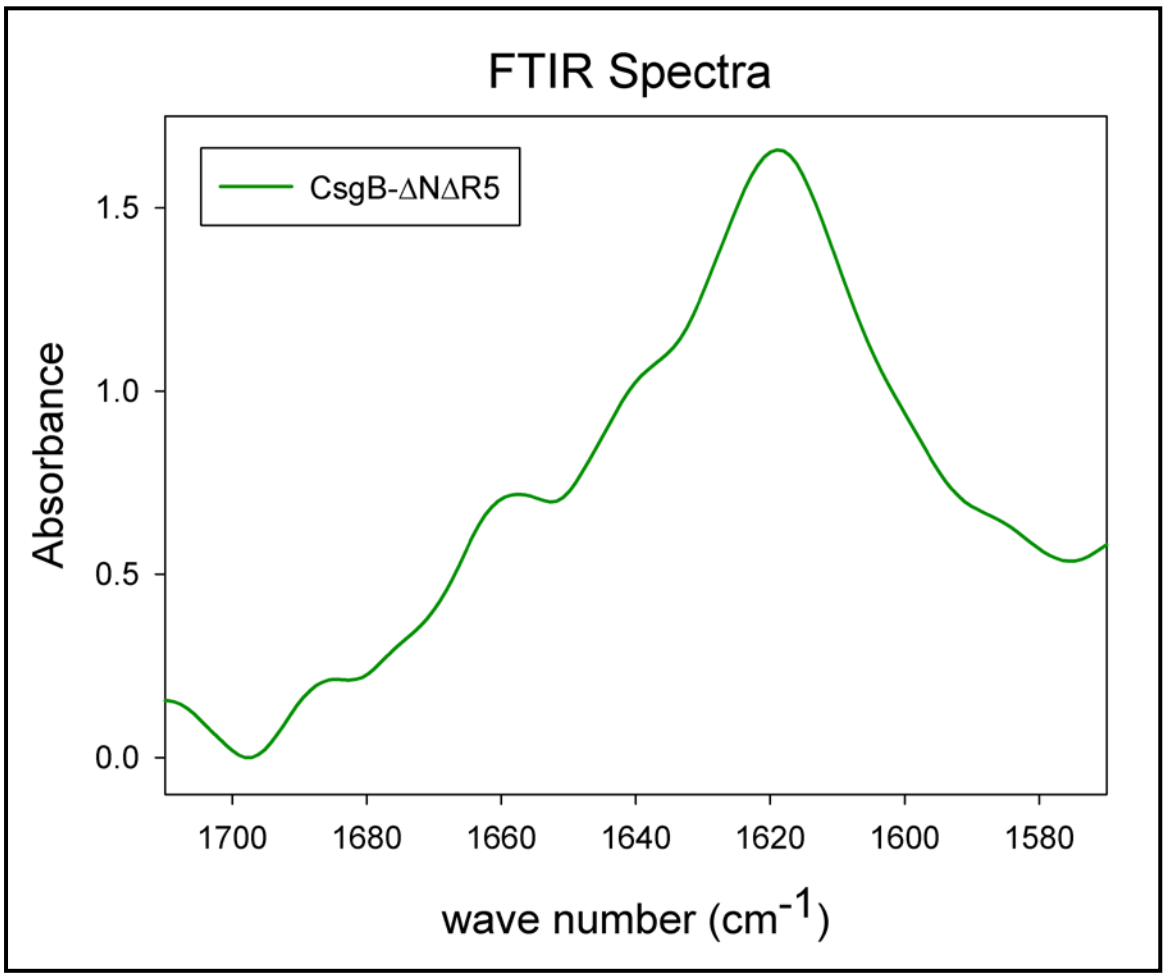
```
 40 SSFNQAAIIG QAGTNNSAQL RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN
 80 DASISQGAYG NTAMIIQKGS GNKANITQYG THHHHHH
```

List of Figures

List of Supplementary Figures

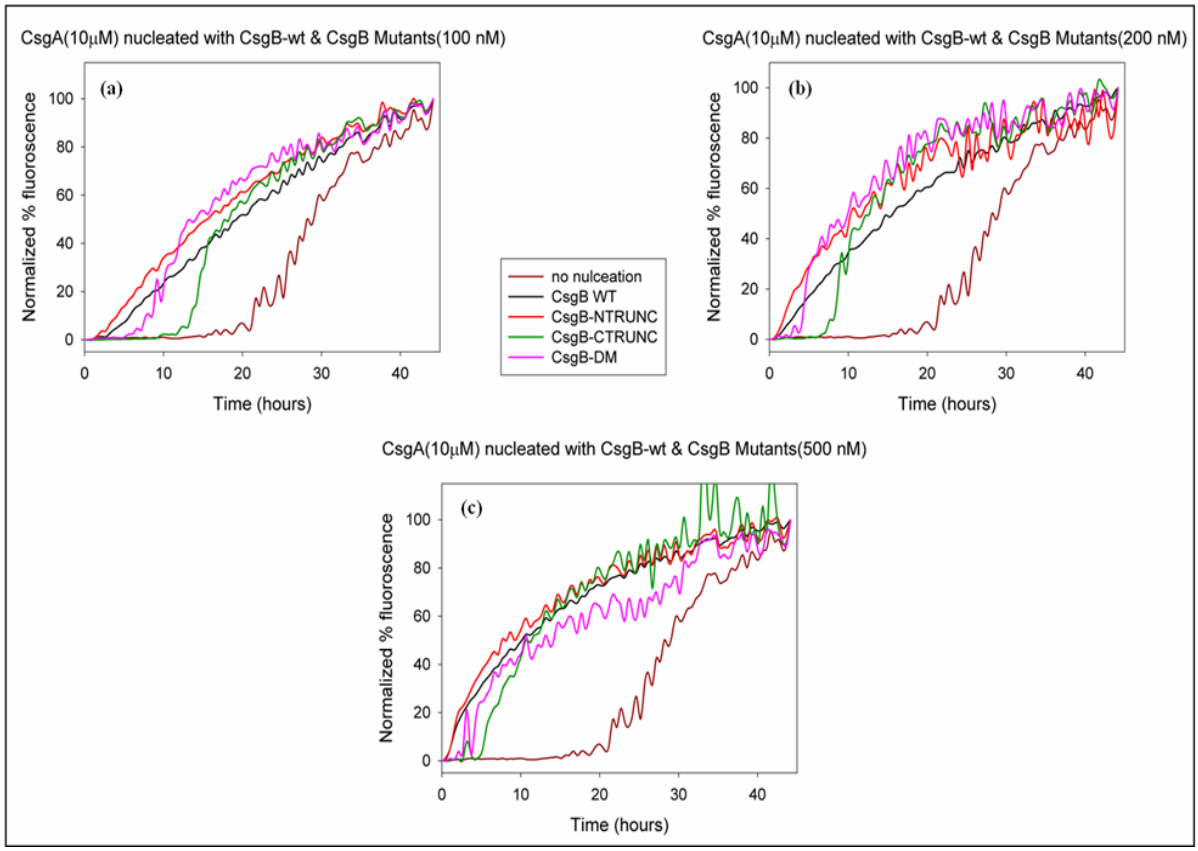
List of Tables

Supplementary Figures



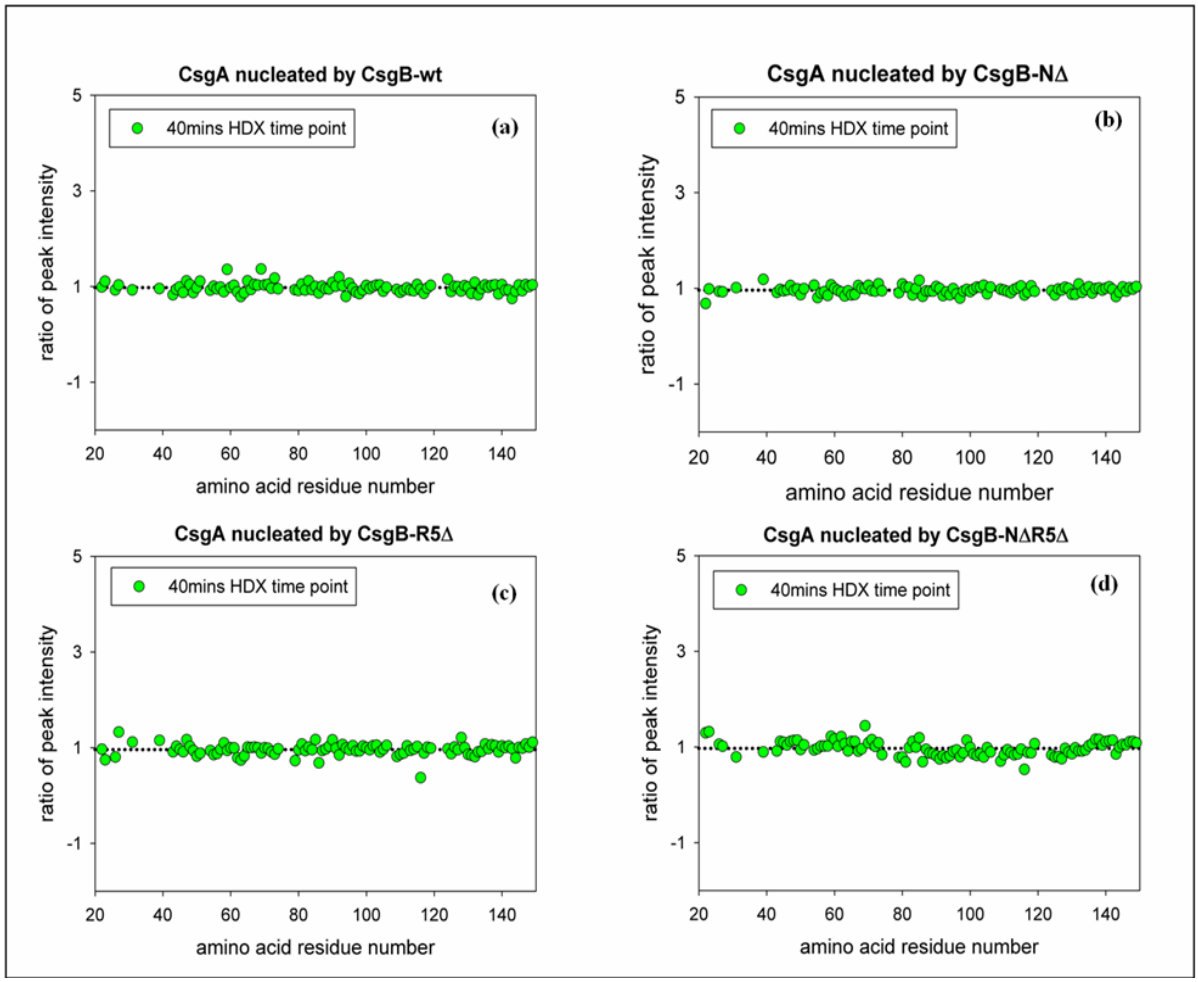
Supplementary Figure 1: Biophysical characterization of CsgB-ΔNΔR5 truncation mutant.

FT-IR spectra showing the amide I band of aggregates (1 mg/ml) formed by CsgB-ΔNΔR5 truncation mutant samples (green curve).



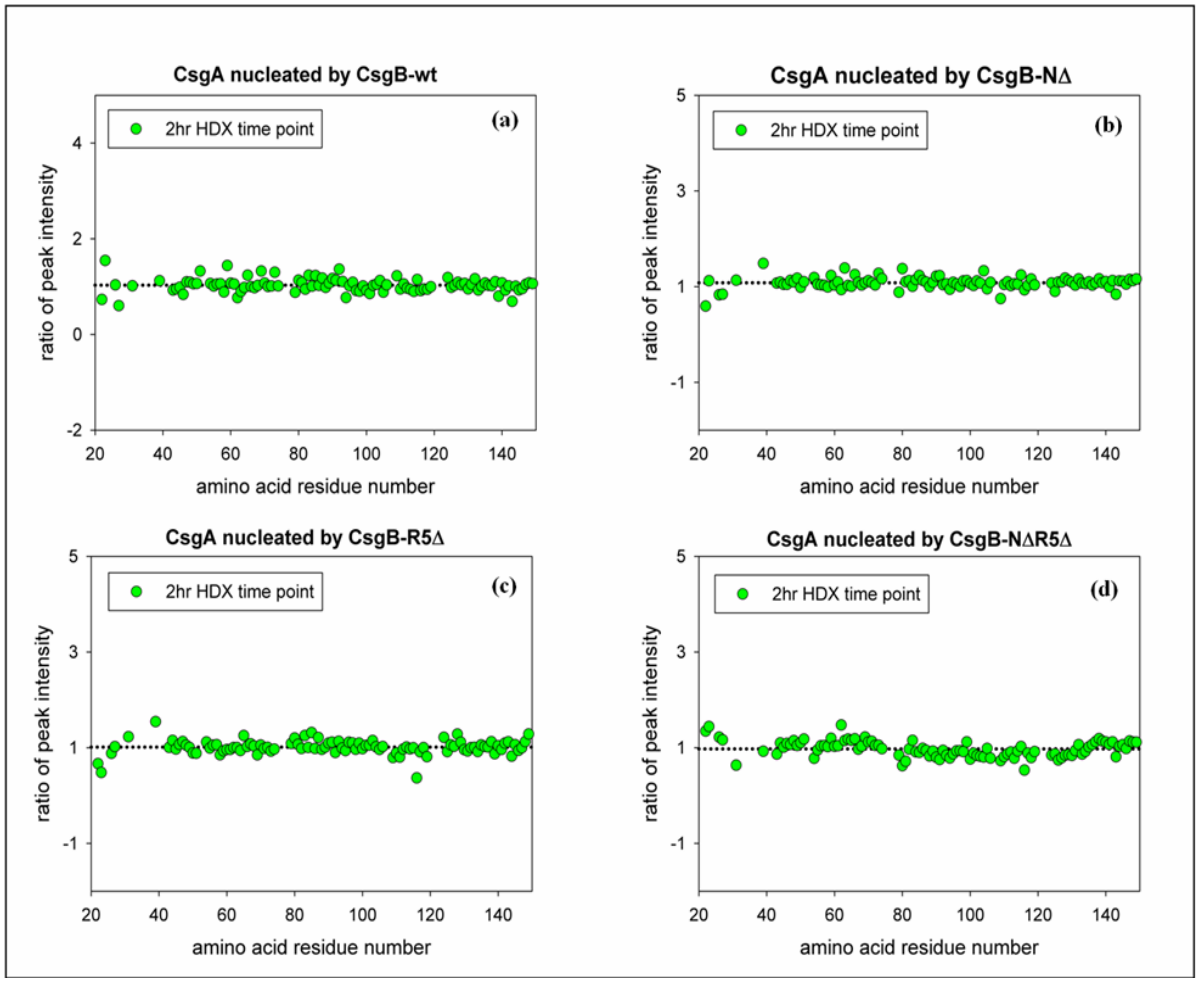
Supplementary Figure 2: Nucleation of CsgA by CsgB-wt and CsgB-wt truncation mutants.

Polymerization of CsgA-wt (10 μM) followed by ThT Thioflavin T in presence of 100nM, 200nM, 500nM monomeric and soluble nucleator CsgB-wt, CsgB-ΔN, CsgB- ΔR5 & CsgB-ΔNΔR5.Nucleation capacity of the CsgB protein follows this order (CsgB-ΔN > CsgB-wt > CsgB-ΔNΔR5 > CsgB-ΔR5)



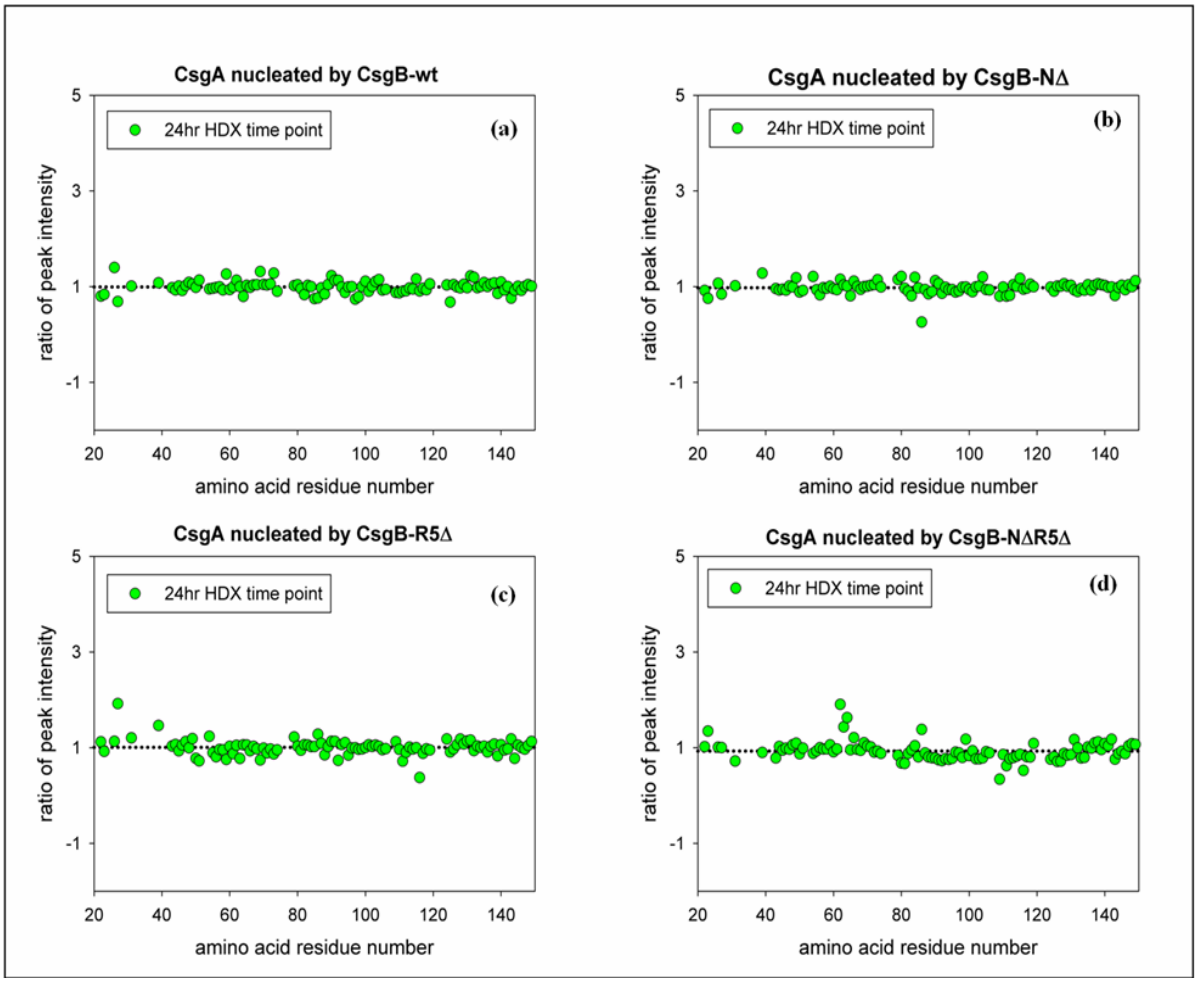
Supplementary Figure 3: Ratio of peak intensity of CsgA fibers to determine polymorphism.

(a) Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-wt over fibers formed by **seeding** with Curli $\mathbf{I_B/ I_c}$. **(b)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔN over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N}/ I_c}$. **(c)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta R5}/ I_c}$ **(d)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔNΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N\Delta R5}/ I_c}$.



Supplementary Figure 4: Ratio of peak intensity of CsgA fibers to determine polymorphism.

(a) Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-wt over fibers formed by **seeding** with Curli $\mathbf{I_B / I_c}$. **(b)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔN over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N} / I_c}$. **(c)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta R5} / I_c}$ **(d)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔNΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N \Delta R5} / I_c}$.



Supplementary Figure 5: Ratio of peak intensity of CsgA fibers to determine polymorphism.

(a) Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-wt over fibers formed by **seeding** with Curli $\mathbf{I_B/ I_c}$. **(b)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔN over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N}/ I_c}$. **(c)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta R5}/ I_c}$ **(d)** Ratio of peak intensity of CsgA fibers formed by formed by **nucleating** with CsgB-ΔNΔR5 over fibers formed by **seeding** with Curli $\mathbf{I_{\Delta N\Delta R5}/ I_c}$.

Supplementary Figure 6: Relative populations in heterogeneous fibers of CsgB-ΔN truncation mutant.

Grey bars indicate the minor population and green indicate major population in case of biphasic behaviour. Residues marked with a star are missing assignments.

Supplementary Figure 7: Relative populations in heterogeneous fibers of CsgB-ΔR5 truncation mutant.

Grey bars indicate the minor population and green indicate major population in case of biphasic behaviour. Residues marked with a star are missing assignments.

Curriculum Vitae

Personal Details

Madhu Nagaraj
Date of birth 25.09.1981 in Bhadravathi, India
Single
Indian

School Education

1987 – 1997 Primary and Secondary School, Bangalore/India

1997 – 1999 Pre-University Education, Bangalore/India

Academic Background

1999 – 2002 Bachelors of Science, Triple major in Chemistry-Botany-Zoology MES College, Bangalore University
Bangalore, India

2002 – 2004 Master of Science, Microbiology
Janabharathi, Bangalore University
Bangalore, India

2005 – 2006 L3 SCIENTIST at JUBILANT BIOSYS (Scientific organization)

2006 – 2008 Junior Research Fellow, Department of Biochemistry
Indian Institute of Science, Bangalore, India

Publications

Jois PS, Madhu N, Rao DN (2008). Role of histidine residues in EcoP15I DNA methyltransferase activity as probed by chemical modification and site-directed mutagenesis. Biochem J. 2008 Mar 15; 410(3):543-53.